

Desperately seeking a syndrome

The US Congress should stop pushing researchers to invent a medical definition for Gulf War syndrome, the collection of maladies associated with veterans of the 1991 conflict in the Persian Gulf.

On frequent occasions in recent years, groups of experts have been appointed to answer the question: is there a Gulf War syndrome? Is there a link between the chemicals, medicines and other materials to which troops were exposed in the conflict and their subsequently reported long-term health problems? The answer, in a recent Institute of Medicine (IoM) study (see *Nature* 407, 121; 2000) and in all previous such excursions on both sides of the Atlantic, has been a polite but firm “no”.

But last week in the US Senate the appropriations subcommittee for labour, education and health, chaired by Senator Arlen Specter (Republican, Pennsylvania), held a hearing on the issue. The IoM study was requested by the Congress to determine, on the basis of an extensive literature review, whether any of the veterans' health problems could reasonably be attributed to their exposure to nerve gas, depleted uranium or vaccines and other medication administered during the conflict. It found that, although some of these exposures could be linked to acute symptoms, there was insufficient evidence to link any of them to chronic health problems.

Specter and Senator Kay Bailey Hutchison (Republican, Texas) used their hearing to advise senior government officials that they should acknowledge what Hutchison termed “the clear common-sense evidence” that the syndrome does exist. But as John Feussner, head of research at the Department of Veterans Affairs, bluntly put it: “The problem with declaring that there is a Gulf War syndrome is that the research suggests that there is not.”

One of the star witnesses at the hearing was Ross Perot, erstwhile presidential candidate of the Reform Party and champion of veterans' causes. Perot has financed the work of Robert Haley, an

epidemiologist at the University of Texas Southwestern Medical Center at Dallas, who believes that he has established several possible links between symptoms reported by a small cohort of veterans and their possible exposure to several different chemicals during the conflict (see *Nature* 385, 187; 1997). The government has supported Haley's work too, to the tune of \$3 million.

Haley argues, justifiably, that the early study of epidemics should concentrate on the close observation of symptoms in small self-selected groups. But Harold Sox, professor of medicine at Dartmouth-Hitchcock Medical Center in New Hampshire and chair of the recent IoM study, told the senators that Haley's studies — which were published in the *Journal of the American Medical Association* almost four years ago — have yet to be independently replicated. Peer-review panels established by the government have rejected Haley's request for a \$25 million follow-up project.

One benefit of the efforts to pin down Gulf War-related illness has been their exposure of the military brass's tendency to downplay the physical and mental impact of war on all of its participants. In future, governments will accept more responsibility for the long-term welfare of returning troops.

In the meantime, there will be no let-up in the political campaign to push for a medical definition of Gulf War syndrome. But as the scant data on who was exposed to what in 1991 fade into the mists of time, scientific grounds for such a definition may well never emerge. The Congress may wish to establish an administrative classification for the health problems afflicting veterans. But it should stop pressing scientists in effect to invent findings that would support its otherwise admirable impulse to assist them. ■

Systems biology's multiple maths

Modelling cellular systems will be a key element of post-genomics science.

Most readers of this publication will know that “post-genomics” and “proteomics” are phrases that mean little that is specific but herald an encyclopaedic era of information about the way biological cells and their genes and proteins behave. But how best to make sense of it all? It is, at last, possible to anticipate mathematics becoming useful in the modelling of the systems. And in that spirit it would be hard to be more ambitious than the efforts of Leroy Hood's Institute for Systems Biology, in Seattle, and Al Gilman's Alliance for Cellular Signalling, based in Dallas.

Hood aims to synthesize gene expression, protein expression and protein interaction into models of specific cell processes, such as immune responses (see page 828). Gilman's consortium aims to characterize every protein–protein interaction in two cell types, then turn the data into a cellular wiring diagram (see *Nature* 407, 7; 2000).

The projects differ both in dimension and scope. Hood's approach, for now at least, straddles several levels of data in a smaller system, whereas Gilman's effort sticks to one level of data in a larger system. These differing emphases will no doubt determine the kinds of models that will emerge from each effort. Gilman's “virtual cell”

would seem best represented by a series of differential equations, as protein–protein interactions are characterized in a linear fashion. Hood's approach appears to demand different kinds of maths — perhaps forms that capture the probabilities of different cellular elements being simultaneously engaged, or algorithms that account for constantly fluctuating cellular conditions.

But such programmes don't come cheap. Gilman's centre has already received \$25 million over five years from the US National Institutes of Health and is seeking an additional \$25 million from private sources. Hood is amassing a \$200 million endowment for his institute. The centres' different approaches complement each other well. But such work is embryonic, and the two centres could well need to exploit each other's developments as each evolves. They could also need to alter their emphasis as more protein and gene information becomes available in the next few years. But provided they and others like them maintain their flexibility to adapt, they should achieve a long-awaited development when mathematics does what it has so often done in other disciplines: provide a basis for prediction and thereby lead rather than lag behind the experiments. ■





Long view

NSF gives ecologists big grants to unravel life's complexity
p823



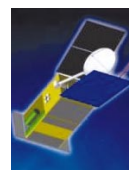
Nazi research

Historians find more skeletons in German science's closet
p824



Going French

Los Alamos preprint archive gets a Gallic outpost
p825



Hot destination

Solar mission joins list of European projects
p827

Tissue donors use their influence in deal over gene patent terms

Paul Smaglik, Washington

A genetic disorder is at the heart of what is believed to be a first for the intellectual property world. A joint patent application for a disease-causing gene is expected to be filed shortly by the researchers who isolated the gene and the patient advocacy group that helped them make the discovery.

The group, PXE International, provided scientists at the University of Hawaii with blood and tissue samples from patients with pseudoxanthoma elasticum, a genetic disorder that causes connective tissue in the skin, eyes and arteries to calcify.

The novel alliance could become a model for handling intellectual property emerging from collaborations between patients' groups and researchers. Sharon Terry, president of PXE International, says that the group made the move because it wants to ensure that licences for any resulting genetic tests will be inexpensive and widely available.

The application comes at a time when both the US National Bioethics Advisory Commission and the Human Genome Organization are emphasizing that genetic researchers should share the benefits of their research with the subjects who make it possible (see *Nature* 403, 237 and 404, 912; 2000).

PXE International filed a provisional patent in February, shortly after the gene was isolated (see *Nature Genetics* 25, 223–227; 2000). It is now working out terms with the university before submitting the final application. The organization advocates free licensing but says that if the university insists on a fee, it will push to keep it low and will make sure any profits are split between the collaborators.

Terry, who calls negotiations between PXE International and the university "amicable", is optimistic that the two will soon come to terms. "Both of us are concerned that tests and treatments be accessible and low cost," Terry says. "We have the same goals in mind."

To protect patients' interests, PXE International set up its own blood and tissue bank. Researchers wanting to use the samples must agree to the group's terms, which include joint possession of any intellectual property that might result.



Right moves: Ian Terry's genetic illness led his mother Sharon to set up PXE International.

Previous collaborations in which the researchers have retained the complete rights to their findings have sometimes raised concerns. One notable example is the isolation of the gene for Canavan's disease, a neurological disorder characterized by the degeneration of the myelin sheath, the protective insulation for the brain's nerve cells.

After supplying samples to the Miami Children's Hospital, many parents were upset when the hospital sought exclusive licensing

on the tests, Judith Tsipis, a biology professor at Brandeis University, who lost a child to the disease, told the American Society of Human Genetics meeting in Philadelphia this month.

Such restrictions are contrary to good science and medical practice, Tsipis said. "Anything short of open licensing limits the abilities of laboratories to do testing." She says that she is pleased that PXE International took control of its members' samples, and hopes more groups do the same.

Elizabeth Thomson, co-director of the Ethical, Legal and Social Implications programme at the National Human Genome Research Institute (NHGRI), says that this is a possibility. Thomson helped PXE International to set up its sample repository and data registry, mostly by giving the group advice on informed consent and drawing up guidelines that would limit how researchers could use the material. The NHGRI did nothing to encourage the group to seek intellectual property, she says.

Other genetic disease groups might follow PXE's example — an umbrella organization for such disease groups, Genetic Alliance, has been watching the group's moves closely. ■

► <http://www.pxe.org>

► <http://www.geneticalliance.org>

Spain's science figures under fire

Xavier Bosch, Barcelona

The science commission of Spain's opposition Socialist Party announced last week that it will file an amendment to the government's proposed science budget for 2001 on the grounds that military spending "has been camouflaged as spending on research and development".

The commission claims that basic research "has been forgotten" in the budget, in which "one in two pesetas are devoted to defence". Its statement follows a report in last week's *Nature* that the government was misleading scientists by including military funds in the budget for research and development (see *Nature* 407, 664; 2000).

Alfredo Pérez-Rubalcaba, minister of education and science in a former socialist government, says that the *Nature* story matches his own data. He claims that only Pta47 billion (US\$250 million) of the total budget increase of Pta572 billion will find its way to basic research. The results would be "a theoretical 6.6% increase — but this is only one point above the growth of the Spanish economy".

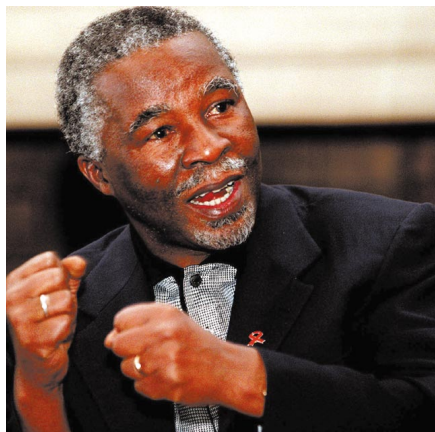
In contrast, the budget for military projects will increase by 12%. Pérez-Rubalcaba emphasized that, although his party is not opposed to investment in defence, "it must not be registered as R&D spending". ■

Mbeki agrees to step back from AIDS debate

Michael Cherry, Cape Town

AIDS researchers in South Africa have cautiously welcomed reports that President Thabo Mbeki has told the national executive committee of the African National Congress (ANC) that he is withdrawing from the debate on the science of the relation between HIV and AIDS.

A party spokesman confirmed that the



Mbeki: faces fears that his stand on HIV/AIDS has led to disillusionment with his leadership.

cabinet has agreed that deputy president Jacob Zuma should chair a cabinet committee on the debate initiated by Mbeki on the link between HIV and AIDS, and the safety of antiretroviral drugs.

According to the *Johannesburg Sunday Times*, Mbeki has told the national executive committee — the party's highest-ranking decision-making body — that his participation in the debate was causing confusion. Party insiders reported concern in the government that the controversy was creating a negative mood in the country and had led to disillusionment with the president. It has also divided the ANC and its allies, the Congress of South African Trade Unions, and the South African Communist Party.

Mbeki apparently told the committee that he would leave his ministers, led by health minister Manto Tshabalala-Msimang, to liaise with the international advisory panel that Mbeki set up to examine the HIV/AIDS issue. He also said that the committee of ministers, not him, would receive and process its report.

The report on the panel's deliberations has yet to be released, although it is said to have been submitted to the directors-general

of health, and arts, culture, science and technology. But it has already met disapproval from both the dissident members of the panel, who regard it as unsupportive of their position, and the orthodox members, who feel it lacks a coherent synthesis of the debate.

A second report, put together by a task-force appointed by the panel to investigate the reliability of HIV testing, should be ready by the end of the year, according to Tshabalala-Msimang. This will report on an experiment in which tests from each of South Africa's five testing centres for HIV will be repeated blind by the Centers for Disease Control and Prevention in Atlanta, Georgia, in an effort to confirm their reliability (see *Nature* 406, 113; 2000).

There has been speculation among AIDS researchers that Mbeki's move may indicate a more pragmatic approach towards the issue. Last week Tshabalala-Msimang announced that HIV-positive pregnant women would be given the anti-HIV drug nevirapine at seven hospitals in KwaZulu-Natal — the country's worst-affected province — in an expansion of the programme to investigate the prevention of mother-to-child infection. ■

UK to make the northwest a post-genomics hotspot

David Adam, London

The future of the Daresbury Laboratory in northwest England, home of the United Kingdom's current synchrotron light source, seems brighter. The centre is one of the biggest winners in a £26 million (US\$38 million) government package designed to enhance the region's science base.

The science minister Lord Sainsbury announced the funding during a visit to the laboratory just outside Manchester on Monday this week. Nine collaborative projects have secured government funding, most of them in genomics and imaging.

"These are areas in which the northwest excels," said Lord Sainsbury. As well as Daresbury, other projects involve researchers at universities and hospitals across the region.

Daresbury staff welcomed the investment, which many will see as a consolation prize after the government's decision to site the country's new synchrotron source, Diamond, at the Rutherford Appleton Laboratory near Oxford (see *Nature* 404, 323; 2000). Synchrotrons generate powerful X-rays used to determine the structure of biological molecules and advanced materials.

"The ground-level feeling around the

place is that the awards are encouraging," says Tony Buckley, spokesman for Daresbury. "But there is also the feeling that this is very much a first course and that dinner is still to be served."

The largest chunk of money, £9.8 million, will help to set up a post-genomics research consortium made up of four local teams, each involving the University of Manchester. Together with the University of Liverpool, Manchester will get £2.39 million to combine physical and biological expertise in a Centre for Bioarray Innovation.

The University of Manchester Institute of Science and Technology joins these two institutions in a £2.01 million Third Generation Proteomics project to develop new ways of probing the protein-protein interactions at the root of many molecular diseases. And, together with Daresbury, the trio will use a £2.14 million grant to work on a Microfluidic Analytical and Screening Technology Centre.

This project sets out to develop high-throughput screening and genome-analysis techniques to support the region's healthcare and pharmaceutical industries. The University of Manchester will go it alone on the fourth post-genomics project: a clinically driven £3.0 million investigation



Jam tomorrow? Staff at Daresbury hope that the new money is a "first course".

of how common diseases develop in individuals and populations.

Other projects include a £2.6 million structural genomics centre to be built at Daresbury, and a £5.75 million multidisciplinary Institute for Functional and Molecular Imaging that will develop and use diagnostic imaging and therapeutic techniques.

"This is great news for the northwest and recognizes the region's international standing in scientific research," said Sir Martin Harris, vice-chancellor of the University of Manchester and chairman of the North West Universities Association. ■

NSF puts big money into complex ecology

Colin Macilwain, Washington

When marine biologist Rita Colwell, newly appointed head of the US National Science Foundation (NSF), announced in 1998 that 'biocomplexity' would be a priority for the agency, many scratched their heads and wondered what the word meant.

The meaning of her commitment, if not the word itself, became clearer last week when the NSF handed out \$52.5 million of grants for multidisciplinary ecological studies.

The 16 main awards, each worth between \$1 million and \$4 million, are remarkably large for university ecologists, even in the United States. Their award represents an important breakthrough for a discipline whose practitioners say they have rarely had the resources they need to do large studies.

"This is going to change our culture," says Diana Wall, an expert in soil biodiversity at Colorado State University, Fort Collins, and former president of the Ecological Society of America (ESA). "For the first time, ecology will be linked with other disciplines to study large and complex environmental problems."

One award will go to Stephen Carpenter of the University of Wisconsin at Madison, the current ESA president. With four others, he will get almost \$3 million to study the interactions of people, land and water on lake shorelines.

Carpenter says that, in preparing the grant application, the team spent much of its time working out just how the project should address complexity. Their eventual proposal was to determine whether it is better to use a complex model for lakeshore ecosystems, or whether a conventional, linear approach will suffice.

Whereas some grants are for studying broad ecosystem questions, others cover specific environmental problems. Alan Hastings and four colleagues at the University of California at Davis, for example, will receive \$4 million to study the impact of the invasive cordgrass *Spartina alterniflora* on America's Pacific coast.

Lest ecologists worry that the money is a flash in the pan, the NSF is also supporting 57 'incubation activities', which will allow groups of researchers with common interests to get to know each other and plan projects for the future.

As the grants were released, Congress agreed on a 2001 budget for the NSF that includes \$75 million for biocomplexity. This is a rise of 50% compared with this year, but is well short Colwell's request for \$136 million (see page 827). Agency officials hope to request proposals for this money as soon as next month.

The biocomplexity initiative is part of a broader push, recommended last year by the



The big picture: a Wisconsin lakeshore, subject of one 'biocomplexity' project.

National Science Board, to increase the NSF's support for environmental science by as much as \$1 billion a year. But the Congress's move to cap biocomplexity's

growth indicates that this is likely to be easier said than done.

► <http://www.nsf.gov/home/crssprgm/bel/biocomp-init/bcawd00.htm>

Deep roots of Nazi science revealed

Alison Abbott & Quirin Schiermeier, Munich

A Berlin-based group of science historians has found that the ties between the top German scientists whose research was used to justify Nazi policies and the regime itself were much closer than was once thought.

In 1939, for example, Heinrich Himmler, the head of the Gestapo, praised two leading researchers at the Kaiser Wilhelm Institute (KWI) for Anthropology in Berlin for their work "that has contributed significantly to the scientific recognition of the racial parts of the national socialist view of the world".

The historians' report was published last week. Their research on the history of the Kaiser Wilhelm Society (KWS) from 1933 to 1945 is being funded by the Max Planck Society (MPS), its postwar successor.

Responding to the report, Hubert Markl, president of the MPS said: "There is no way of denying that directors, scientists and research assistants in many biomedical Kaiser Wilhelm Institutes put themselves in the service of a criminal regime. They actively supported measures conflicting with human rights and exploited the opportunities to use science conducted beyond all morally acceptable limits for the benefit of their own research."

Markl said he felt "deep disgust for the way that unscrupulous research was carried out" and appealed to German biomedical researchers to "confront and discuss their

director of the KWI for Psychiatry in Munich, and Ernst Fischer and Otmar von Verschuer, both of whom headed the KWI for Anthropology in Berlin during the Third Reich, advised the Nazis at the highest levels.

All three had direct contact with the Nazi leadership and served on important government advisory panels. Rüdin, along with his co-worker Fritz Lenz, sat on an expert committee to the ministry of the interior on population and race policies.

Research into the links between KWS scientists and the Nazi regime dates back to the early 1980s. But the Berlin group is the first to investigate the direct influence of contemporary anthropological and biomedical research on the Nazi's racial and medical legislation.

Historians knew that Rüdin, whose work to give race laws a scientific basis was funded directly from Hitler's office, had chaired the committee's working group on 'racial hygiene and racial policies'. This panel set the criteria for the castration of criminals and the forceable sterilization of so-called inferior women, particularly those with 'psychological' problems.

The report says that Rüdin lobbied successfully for ever broader criteria. On his initiative, the sterilizations came to include the 'morally ill' — Nazi-speak for the mentally handicapped. This category covered 95% of the 400,000 sterilizations carried out between 1933 and 1945.

At Rüdin's suggestion, the sterilized

► © 2000 Macmillan Magazines Ltd
The group has found that Ernst Rüdin,

included 600 children of black French soldiers and German women in the state of Rhineland, which the French occupied after the First World War.

In 1937, von Verschuer was made an expert to the Research Department on Jewish Questions of the governmental Reich's Institute for the History of the New Germany, whose main role was to research 'world Jewry'.

In this capacity, he wrote to Fischer that it was "important that our race policies, including the Jewish question, develop an objective scientific background that is broadly accepted".

The historians have not yet been able to show whether Nobel laureate Adolf Butenandt, director of the KWI for Biochemistry in Berlin and a postwar president of the MPS, allowed blood samples from Auschwitz to be analysed at his institute (see *Nature* 403, 474–475; 2000). But they argue that it is unlikely that he did not know about the work being done under his directorship.

The historians have so far focused on four institutes. Three of these were considered biological; the fourth was the KWI for Metal Research in Stuttgart.

It is not known whether Nazi sympathies were the exception or the rule among scientists, says Carola Sachse, the research group's head. ■



Close ties: Max Planck (centre) at the Kaiser Wilhelm Institute for Metal Research in 1935.

US court slashes damages in polymerase- β theft case

Rex Dalton, San Diego

How much is academic research worth? For lab work that took nearly three years and produced papers in *Science* and *Biochemistry* the value is one US dollar, according to a California appeals court.

The ruling covers a lengthy battle over research on the rat DNA enzyme polymerase- β , which was taken by La Jolla-based drugs company Agouron Pharmaceuticals from a federally funded laboratory at the University of California at San Diego (UCSD).

Last month, the state appeals court in San Diego upheld a 1998 jury verdict that in 1992 Michelle McTigue and her husband Jay Davies, both of Agouron, stole key elements of crystallography research from Huguette Pelletier, then a postdoc at UCSD (see *Nature* 393, 504; 1998). But the court reduced the jury's \$200,000 award to Pelletier to \$1, ruling she had not proved any damage.

Formerly a crystallographer at Baylor College of Medicine in Houston, Texas, Pelletier left science last year, largely in frustration over the incident. She now works for a Los Angeles community-service organization.

The decision highlights the dilemma faced by university scientists trying to protect their research from the representatives of biotech companies who walk round university labs seeking leads to develop into products.

Pelletier will appeal to the California Supreme Court against last month's verdict. She hopes to create a legal precedent affirming the value of academic research.

She has written to the Bruce Alberts, the president of the National Academy of Sciences, asking him to file documents with the California Supreme Court on the value of academic research and the damage caused by scientific misconduct.



Pelletier: will appeal against the verdict.

"It is difficult under existing laws to place a dollar value" on research such as Pelletier's, says her attorney, Daniel MacLeod. The \$200,000 originally awarded was based on a year's grant from the National Institutes of Health, which funded the lab where she worked at UCSD.

Academic scientists who are victims of this sort of practice have to seek redress under laws designed

for misappropriation of commercial and patented data, MacLeod says. And as the reward system for academic researchers involves a range of factors, including publication and academic promotion, he adds, it is difficult to demonstrate damage.

Mike Varney, vice-president for research at Agouron, now owned by the drugs company Pfizer, issued a statement saying that the company "continue[s] to fully support the integrity of our scientists and the appropriateness of their conduct."

Pelletier filed her lawsuit in 1994. She contended that McTigue, a former doctoral student at UCSD, used her access to the university to secure techniques on crystallography on rat polymerase- β . These were then given to Davies for his Agouron project to try to develop a cancer medication from the enzyme, which repairs DNA.

The project was eventually abandoned. But Davies, McTigue and others used the data for the first publication of the rat polymerase- β crystal, which appeared in *Cell* in 1994. Pelletier's work on polymerase- β was later published in *Science* and *Biochemistry*. ■

Canada plans reduction in greenhouse-gas emissions

David Spurgeon, Montreal

A month before the next large international meeting on climate change, to take place in The Hague in the Netherlands, the Canadian government has announced that it will cut emissions of greenhouse gases by about 65 million tonnes per year over the next five years. This would take it one-third of the way towards the target agreed at the previous climate meeting in Kyoto.

The government will spend up to Can\$500 million (US\$330 million) on cutting emissions from transport, energy

production, industry, buildings, forestry and agriculture, international projects, technology and science. Together these activities account for more than 90% of the total.

Since 1995, Canada has spent Can\$850 million on climate change, but its greenhouse-gas emissions have continued to rise. An Environment Canada report found that they were 13% higher in 1998 than in 1990.

Canada's Kyoto target is to cut the 1990 figure by 6% by 2008–12. According to the government figures, this will mean a

reduction of some 200 million tonnes.

Canada's "Action Plan 2000 on Climate Change" promises that the government will buy as much as 20% of its electricity from renewable sources such as wind and solar power. It also includes incentives to stimulate sales of renewable energy and a tripling of the amount of ethanol the country produces annually.

The plan will be taken to a meeting with the provinces later this month in Quebec City in an attempt to convince them to make similar commitments in advance of the meeting in The Hague. ■

French take physics archives into the future

Declan Butler, Paris

France's national research agency, the Centre National de la Recherche Scientifique (CNRS), is to establish a French arm of the Los Alamos preprint archives of physics papers. The move is part of wider plans being discussed by the International Union of Pure and Applied Physics (IUPAP) to upgrade and internationalize the archives.

The French arm, the Centre pour la Communication Scientifique Directe, will be based in Lyons and will consist of a handful of full-time computer scientists and engineers. Physicist Franck Laloe, who will head the unit, says it will be much more than a mirror site, aiming to develop the archives and extend them to other disciplines.

An international steering committee will meet later this month to advise the centre; it includes Paul Ginsparg, the founder of the Los Alamos archives, and Martin Blume, editor in chief of the American Physical Society.

"It's marvellous that CNRS are forward-looking enough to try to forge the first formal elements of an international collaboration, and also to move beyond the primary physics and math components of the existing



Ginsparg: backs international collaboration.

archives into other fields," says Ginsparg. "Ultimately, one could imagine every research institution and/or every funding agency joining in to help create the underlying data network."

Ginsparg has long argued that research institutions and their funding agencies should take greater control of the publication of raw research results.

"The costs of publishing on the Internet are a tiny fraction of what institutions pay in salaries and overheads for the research itself, and in this way they can ensure that the primary literature is made freely available," he says, adding that the cost of maintaining the

Los Alamos archives is as little as one dollar per article.

Blume says he is "delighted" about the CNRS decision. "At the meeting, I will encourage discussion about the establishment of an international consortium to support Paul Ginsparg in the operation and funding of his archive."

The IUPAP working group on communication in physics, which Blume chairs, last month recommended that the union should "play a role in assuring that there is international participation" in funding the archives, assuring availability and, if possible, aiding their operation "by mobilizing physics societies worldwide".

One drawback of the archives at present, says Blume, is that submissions can only be made through Los Alamos. To improve the situation, software needs to be developed that will allow submissions to be made at any one of the several mirror sites worldwide, he adds.

The American Physical Society (APS) initially viewed the archives as a competitor, but it is now actively collaborating. Since last year it has mirrored the archives, and uses them as an electronic submission procedure for APS journals. ■

Anger as Princeton closes 'inspirational' museum

Rex Dalton

Princeton University closed its Natural History Museum in Guyot Hall last month to provide space for a new environmental research unit. The move has angered university graduates and researchers worried that its historic collection may disappear into storage or be damaged when moved to a promised new facility.

Opened in 1909, the museum on the New Jersey campus drew large numbers of adults and children each year. It housed specimens from a century of palaeontological and geological discoveries by explorers from the university, including artefacts from a prehistoric Swiss village discovered in 1878, specimens of the moa — an extinct giant bird — and a virtually complete *Allosaurus*.

Princeton officials say they plan to house some specimens in a \$65 million extension, to be built in the next five years, which may combine biological, atmospheric and Earth sciences. But some current and former members of the geosciences department fear the collection could be broken up and distributed to other facilities, severing Princeton's ties to a golden era of discovery.

In the early 1980s, Princeton gave its

vertebrate palaeontology research collection to Yale University, a move that prompted a similar swell of discontent. "The university is making a serious error in getting rid of this inspirational space," geology professor Lincoln Hollister wrote to Princeton president Harold Shapiro.

Hollister is among a group calling for a year's moratorium on dismantling the museum so funds can be raised to enhance it. Shapiro was unavailable for comment.

Meanwhile, the space is being converted into offices for the Environmental Institute, the recipient of major research grants from the National Science Foundation and British Petroleum, which is sponsoring research into carbon sequestration in old wells.

"I think it is deplorable that Princeton is closing the museum," says Lydia Fox, a Princeton graduate who is now chair of geosciences at the University of the Pacific in California. "A natural history museum can be a critical form of outreach to an increasingly science-illiterate citizenry."

More than 100 geoscience graduates have asked Princeton's president to retrieve the museum. Last month, a study showed the specimens were extremely valuable, but vulnerable to damage if moved.



Oceanographer George Philander, Princeton's chairman of geosciences, is apologetic for the way that the university has handled the museum closure, saying he "completely underestimated the emotional attachment" to the facility.

But Philander insists that the planned new facility will be popular. Museums like the one in Guyot Hall "have been able to describe what the world was like," he says. "Now we are on the verge of being able to describe why the world is the way it is. We want a museum to reflect that." ■

Genetic screen gets go-ahead in UK life insurance premiums

London In what is believed to be the first decision of its kind, British insurance companies have been given the go-ahead by a government advisory committee to use the results of genetic screening tests — in this case for Huntington's disease — in the assessment of life insurance premiums. One of the more common genetic conditions, Huntington's disease is a degenerative brain disorder.

Last week, the Genetics and Insurance Committee of the UK Department of Health announced that it considered the current tests for the disease to be reliable. It added that permitting the use of genetic test results would mean that those individuals with a family history of Huntington's disease, but who tested negative, would not be asked to pay more for life insurance.

The move has been criticized by the National Consumer Council, which claims that the fear of not being granted insurance could dissuade individuals from taking the genetic tests. But it was welcomed by the Association of British Insurers, whose members are awaiting further decisions from

the panel on six other screening tests.

♦ <http://www.doh.gov.uk/genetics/gaic.htm>

Australian biologists focus on crop maturity genes

Sydney Plant molecular biologists in Australia believe they will be able to isolate a gene that regulates the maturing of cereal crops. The claim follows their recent success in identifying Flowering Locus C, a gene that determines the time of flowering in the weed *Arabidopsis* and in the oil seed canola.

Liz Dennis, a chemist with Australia's Commonwealth Scientific and Industrial Research Organisation (CSIRO), predicts that control of flowering could add up to 10% to wheat productivity, and expects this to be possible within six years. Last week, Dennis and her colleague Jim Peacock were recognized with the Australian Prime Minister's Prize for Science (formerly known as the Australia Prize), worth A\$300,000 (US\$159,000).

Russia and India forge stronger science links

New Delhi President Vladimir Putin of Russia has signed a ten-year science agreement extending the Integrated Long Term Programme (ILTP) instigated by Soviet

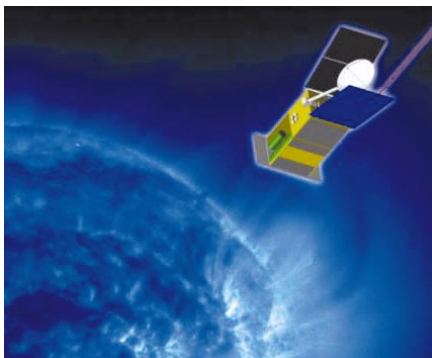
leader Mikhail Gorbachev and Indian President Rajiv Gandhi in 1987. Visiting New Delhi two weeks ago, Putin also promised more Russian help for India's nuclear power and space programmes.

The ILTP has generated over 300 research projects. It has also led to the establishment of research centres in India on new materials and the production of oral polio vaccine, and in Moscow for computing research. Indian officials say that one outcome of Putin's visit is Russia's apparent readiness to divert a small percentage of the several million dollars earned from selling military hardware to India to the cash-starved ILTP.

Europe confirms planned mission to Mercury

Munich The Bepi Colombo mission to Mercury will, as expected, be the next cornerstone mission for the European Space Agency (ESA), with a budget of around 600 million euros (US\$514 million) and a planned launch date of 2009. After that will come GAIA, an astrometry mission to be launched before 2012. The ESA, following the recommendations of its space science advisory committee, formally approved both projects last week (see *Nature* 407, 277; 2000).

The agency also confirmed that it will participate in LISA, the NASA-led mission to



In the pipeline: the European Space Agency's Solar Orbiter is likely to be launched in 2007.

search for gravity waves, and the Next Generation Space Telescope. The Eddington mission to study stellar evolution and search for Earth-like planets orbiting other stars is being held in reserve in case the telescope or the planned Solar Orbiter stumble.

Finland takes the chair in EUROHORCs

Helsinki Reijo Vihko, president and director general of the Academy of Finland, has been elected to chair EUROHORCs (European Union Research Organisations Heads of Research Councils) for the next two years. The election took place during a plenary

meeting of the organization held in Inter-laken, Switzerland.

EUROHORCs, which was set up in 1992, currently brings together the heads of 33 research organizations from all EU member states. The research councils of Norway and Switzerland attend with associate member status. Vihko says that a major challenge facing the organization is how to develop further the collaboration of funding among its member organizations.

MRC launches scheme to boost scientific high-fliers

London Britain's Medical Research Council (MRC) is investing £1 million (US\$1.46 million) in a new predoctoral fellowship scheme intended to boost the number of talented students going into scientific research careers. Beginning in October 2001, 15 students a year will be accepted to study for PhDs for three years while working as salaried employees in MRC institutes and units.

Students on the scheme will receive around £17,000 per year, almost double a typical postgraduate annual grant. According to the MRC, they will also enjoy "greater entry-level research career opportunities". The scheme is one of a number that have recently been set up in the United Kingdom to stem the loss of high-

calibre science graduates into alternative, better-paid careers.

Biggest-ever rise for NSF budget next year

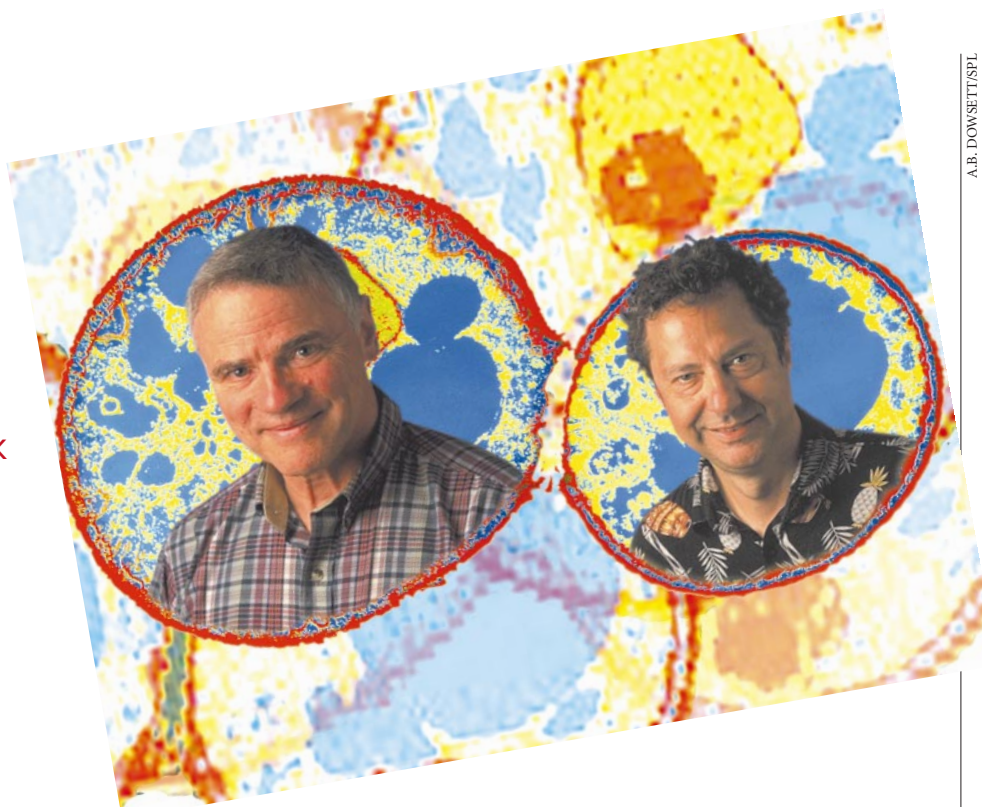
Washington The US Congress has approved the largest-ever budget increase for the National Science Foundation (NSF), increasing its budget for next year by \$526 million to more than \$4.4 billion. Although less than the 17% increase proposed by President Bill Clinton, the bill has thrilled agency supporters, who say that it puts the NSF on track to double its budget over five years.

Agreed by both houses of Congress last week, the appropriations bill for the agency provides money for some of the president's special projects at the NSF. It spends \$75 million on biocomplexity (see page 823), against a \$136 million request, \$150 million on nanotechnology (against \$217 million requested) and \$215 million on information technology (against \$327 million).

Correction A brief news item on the campaign promises of US presidential candidate George W. Bush (*Nature* **407**, 438; 2000) was mistakenly accompanied by a photograph of his father, George Bush, who was campaigning on his behalf in the same state on the same day. We apologize for the error.

For my next trick...

Is it possible to produce a complete mathematical description of complex biological systems? Leroy Hood thinks so, as Paul Smaglik discovers.



A.B. DOWSETT/ISL

Many people in Leroy Hood's position would be resting on their laurels. In the 1980s, while at the California Institute of Technology in Pasadena, he led a team that invented the automated gene sequencer — spawning a multimillion dollar business and helping to lay the foundations for the Human Genome Project. In 1992, believing interdisciplinarity to be the key to progress in biology, Hood convinced Microsoft founder Bill Gates to establish the Department of Molecular Biotechnology at the University of Washington in Seattle. The department united biologists with mathematicians, chemists, engineers, applied physicists and computer scientists, and went from strength to strength — boasting an annual budget of US\$35 million by 1998.

But in December last year, Hood resigned from his academic post to take on his biggest challenge yet. The university, facing a long list of requests for building projects, had declined to back Hood's plan to create an institute for the emerging discipline of systems biology — so Hood decided to strike out on his own. He founded the independent Institute for Systems Biology (ISB), a non-profit research corporation housed in some 3,000 square metres of office space near the university campus. Here Hood is now assembling a multidisciplinary team to meld biology with the tools of mathematics to provide a complete description of biological processes and systems. "Biology is an information science" is his mantra.

Hood has already been joined by four of his former university's brightest stars — scientists whose expertise spans proteomics, immunology and cell biology, as well as the

application of computer algorithms to some of the thorniest problems in astrophysics. Their goal is to give molecular biology a fresh outlook. "The approach to biology for the past 30 years has been to study individual proteins and genes in isolation," says Hood. "The future will be the study of the genes and proteins of organisms in the context of their informational pathways or networks."

A model approach

Hood is not alone in embracing this systems approach. For example, a consortium called the Alliance for Cellular Signalling, which is based at the University of Texas Southwestern Medical Center at Dallas and headed by Nobel laureate Al Gilman, aims to compile a map of the interactions between 1,000 proteins in two types of mouse cell. But Hood's vision is more expansive. In addition to studying simple systems in model cells, the ISB plans to study the immune system and complex conditions such as cancer and heart disease.

To finance these projects, Hood wants to supplement grant funding by seeking endowments for the ISB worth some US\$200 million. He aims to swell the institute from its current roster of some 80 staff to around 400. An initial US\$5 million came from an anonymous donor, and in July the drugs giant Merck & Co. provided a matching sum. Negotiations are underway with further corporate donors, and other companies are providing contributions in kind. The agribiotech firm Monsanto, for instance, has let Hood keep the 30 automated DNA sequencers he used when leading the team that in April completed a draft of the rice genome.

Under the hood: astrophysicist George Lake (right) will help Leroy Hood assemble the biological equivalent of a virtual car.

But even with the finance in place, the road ahead will be difficult. Hood likens the data gathered by the Human Genome Project and other genomics efforts to an incomplete parts list for a car, in which some entries are missing and others are illegible. Although some parts have been grouped with others and seem to fit together, others remain unidentified and isolated. The ISB aims to build from this information and produce the biological equivalent of a virtual automobile.

Car trouble

Developing the mathematical models on which this effort will depend is much more complicated than assembling a car's engine from a pile of incompletely labelled parts. For a start, biological systems encompass several different levels of information. On one level, a cell can be characterized by recording the genes that are expressed at any one time. But to understand the cell as an integrated biological system, you also need to study the structures of the proteins the genes encode, and the interactions of these proteins with each other as well as with other genes. "At the higher levels, things happen that you can't predict at the lower levels," says Hood. Consider also that each gene can code for more than one protein and that several genes can work together to produce individual proteins, and the enormity of the task facing the ISB becomes clear.

Trying to render each element's role in a process mathematically is also extremely

hard. Some systems biologists aim to characterize every aspect in a system as fully as possible, then construct the cellular equivalent of a wiring diagram — using differential equations to describe the interactions between each node. Gilman's consortium, for example, appears to be heading initially in this direction. Others believe that, because cellular systems are constantly changing, such a linear approach will fall short of providing a full description of a biological system. Instead, they are putting their faith in analysing how different components, such as messenger RNA (mRNA) levels and protein expression, relate to each other under various conditions, and then using probabilities to recreate these relationships. The idea is that this better captures the dynamic, non-linear workings of a biological system.

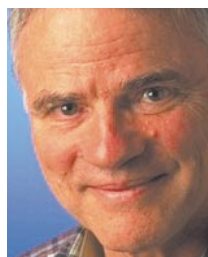
Both approaches may find their niche. But either way, Lee Hartwell, president and director of the Fred Hutchinson Cancer Research Center in Seattle, and a member of the ISB's scientific advisory board, believes that the key is to keep things simple, at least to start with. If scientists at the institute want to characterize every aspect of every system they are interested in, "then they're way too early", he says. Instead, Hartwell recommends a modular approach, concentrating on understanding each component of a system, then integrating them later.

Sweet success?

Hood seems to be heeding that advice. Returning to his automotive analogy, the institute will first produce schematics showing how parts fit together into discrete systems — the transmission, for example — and will then build mathematical models that predict what these component systems do when they work in concert. For now, the ISB is working with relatively well-understood and simple systems in model organisms — akin to understanding how the brakes work in a vintage car.

Initial studies, so far unpublished, have been on sugar metabolism in yeast. As the yeast genome has been sequenced and sugar metabolism in the organism is fairly well characterized, Hood decided to use it as a benchmark to see if systems biology could point towards the same conclusions. ISB scientists, including postdoc Vesteinn Thorsen and graduate student Trey Ideker, performed experiments on yeast fed a steady diet of either the sugar raffinose or galactose. They focused on a handful of genes known to help yeast process galactose rather than raffinose, knocking them out one at a time. After each knockout, the researchers used microarrays to measure which other genes were expressed, and used a novel mass spectrometry technique to measure the abundance of individual proteins.

Measuring both gene expression and protein production paints a clearer picture, says



The future will be the study of the genes and proteins of organisms in the context of their informational pathways or networks.

Leroy Hood



Systems analysts: in addition to Hood and Lake, the Institute for Systems Biology's staff includes (from left) proteomics expert Ruedi Aebersold, immunologist and cell biologist Alan Aderem, and Ger van den Eng, who works on high-speed cell sorting.

Ruedi Aebersold, the ISB's proteomics expert, and one of its high-profile recruits from the University of Washington. Just measuring levels of mRNA, as is typical in gene-expression studies, tells scientists that a gene has been activated, but does not detail the amount of protein it encodes, or whether that protein is functional. The automotive equivalent, explains Aebersold, is knowing that an engine contains a spark plug. The protein analysis, he says, allows you to ask: "Is this spark plug actually firing or is it just sitting in the engine doing nothing?"

The scientists detected around 1,000 genes involved with the switch to galactose metabolism. They then used a series of algorithms to group the mRNAs produced by these genes into families, based on common patterns of changes in the knockout experiments. Finally, they integrated the mRNA and protein data into a simple mathematical model that illustrates how yeast adapts to use galactose rather than other sugars.

Comparing the results with what is actually known about yeast sugar metabolism revealed that the model still needs work. "It actually does a reasonable job of predicting a lot of things, but it's way, way too simple," says Hood. The ISB's scientists will now tweak their algorithms, do more experiments, then adjust the model some more. "The whole process is incredibly iterative," says Hood.

Developing better computer algorithms will help. And the ISB has just gained a valuable asset in the form of another recruit

from the University of Washington: George Lake, an astrophysicist who is project scientist for the Earth and Space Sciences part of NASA's High Performance Computing and Communications initiative. Lake's forte is developing algorithms that, until now, have mostly been applied to problems of cosmology, astrophysics and planetary science. His algorithms, for instance, have made simulations of planetary formation run a million times faster.

Lake believes that building the models Hood needs will be tricky, but not impossible. He likens the challenge to modelling global climate, another field to which he has applied his skills. Climate models have improved through constant mathematical tinkering and by incorporating new kinds of data, says Lake, and the same will be true for simulations of biological systems. He is enthused by his crash course in biology. "It's scary and daunting," he says. "Quite frankly, it just wasn't scary and daunting to go to the office as an astrophysicist anymore."

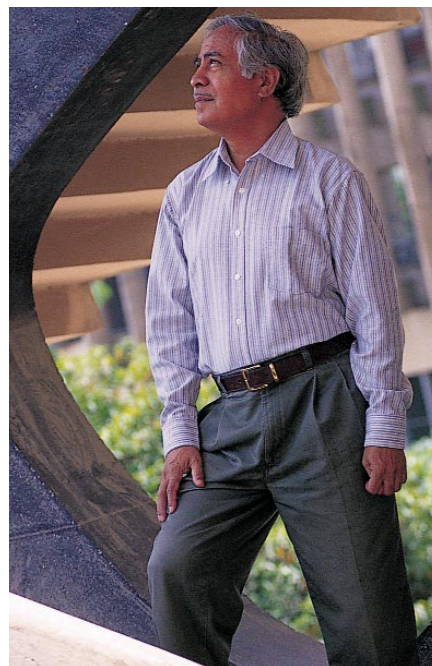
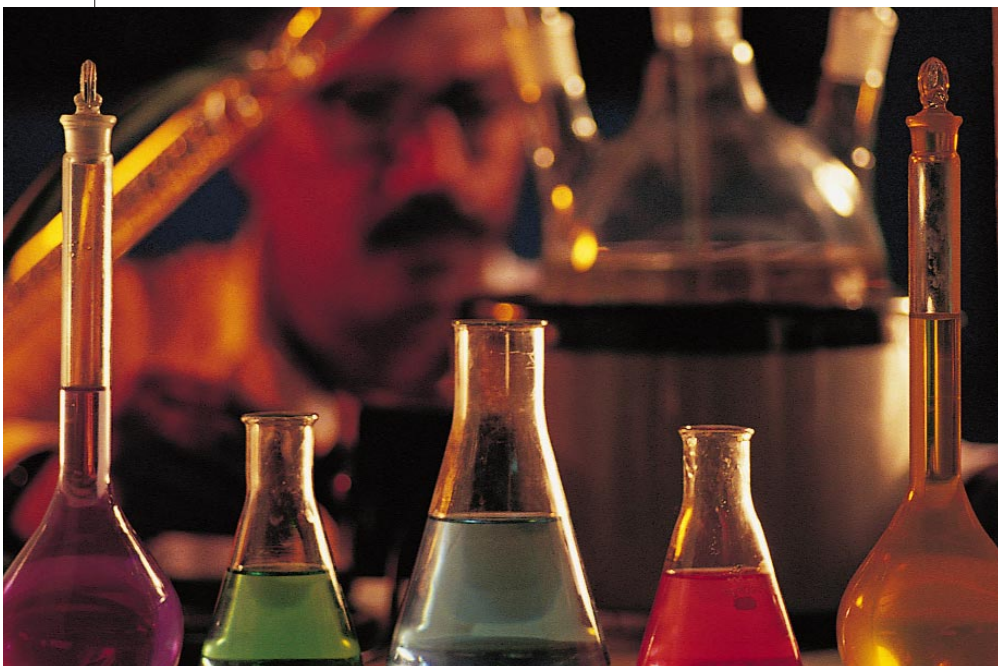
Hood's expertise in technology development might also prove to be important when it comes to gathering the new kinds of cellular data that Lake predicts will be needed. It is still too early to tell whether systems biology's virtual automobile will ever move with the speed and grace of a sports car. But given Hood's track record, if it does, it is a fair bet that the ISB will help lead the way. ■

Paul Smaglik is *Nature's* Washington DC correspondent.

♦ www.systemsbiology.org

India's finest, for hire

By collaborating with western companies, are India's research institutes consolidating their positions or allowing their young researchers to be exploited as a cheap scientific labour force? K. S. Jayaraman investigates.



DINESH KHANNA

On entering the campus of the Indian Institute of Technology (IIT) in New Delhi, the very first building you see carries the logo of IBM. When IBM's New Delhi lab was set up three years ago, Paul Horn, the company's worldwide head of research, promised that this lab would "become a world class facility", with scientists from IBM and the IIT collaborating to develop "software solutions for challenging problems in information technology".

But since then, the IIT faculty's relationship with IBM has gone sour — as it has with several other western companies that have established a presence on the IIT's campus. In fact, last November, the majority of the IIT's faculty backed a resolution to prevent commercial organizations opening premises within the institute. The resolution also said: "Steps should be initiated to nullify the existing agreements."

This vote reflects a wider concern felt by many Indian academics. They believe that India's national research priorities are being distorted by links with western industry. And they are particularly worried that the country's highly qualified young researchers — its most valuable scientific asset — are being used as cheap labour to address the problems of multinational companies, rather than the issues facing India's developing economy.

Given its high profile, IBM's New Delhi

lab has been a particular focus for criticism. Many IIT faculty members complain that the company has failed to fulfil its promises of a mutually rewarding, collaborative relationship. "IBM set up shop here because the IIT is the best address in the city and its students are available to provide cheap labour," claims H. B. Mathur, a former professor of mechanical engineering at the institute. "IBM can stay in the campus provided they want to work with us," adds J. S. Rao, another mechanical engineer. Rao felt so strongly about the issue that in July 1999 he resigned as faculty representative on the IIT's board of governors. "If they want to remain an island, they can step outside," Rao says.

Corporate collaborations

Links between Indian research labs and western companies began to grow in the early 1990s. The trigger was a government policy putting pressure on research institutions to generate their own money to supplement public funding. Multinational companies were only too happy to oblige, paying to set up new facilities on campuses or donating equipment to individual labs. The lure, in part, was India's able and motivated, yet relatively cheap, scientific labour force. Just as India's software engineers are in demand, either working in India on contract to western companies, or being lured

to work in high-tech centres such as Silicon Valley, its brightest young scientists are attractive to western companies.

The IIT constructed a new building on a prime plot within its campus and invited IBM to set up its lab there, charging an annual rent of about US\$450,000. Similarly, the agribiotech giant Monsanto set up a research centre at the Indian Institute of Science in Bangalore, agreeing to pay US\$300,000 per year. And Germany's Hoechst gave US\$150,000 to the National Institute of Immunology in New Delhi in exchange for first rights to discoveries that the institute had made in the fields of autoimmune disorders and drug delivery.

Also during the 1990s, labs run by the Council of Scientific and Industrial Research (CSIR), India's largest scientific agency, began taking up contracts from multinational companies including SmithKline Beecham, DuPont and General Electric. The CSIR's total earnings from contract research reached US\$7.5 million last year, about a tenth of what it receives in government funding.

Proponents of these moves say that the influx of money provides jobs for scientists who might otherwise be unemployed. "The quality of research has also improved and, thanks to the training component of the agreements, our scientists have learned to run instead of crawling," says the CSIR's



Sell out? IBM's research facility (left) at the Indian Institute of Technology in New Delhi (above) has caused controversy. Researchers such as J. S. Rao (opposite) complain that the company has failed to work with the faculty.

director general, Ragunath Mashelkar, who took up the post five years ago with the promise that the agency would become financially self-reliant. Contract research, he argues, does not "conflict with our own priorities as it involves less than 2% of our scientists in seven out of our 40 laboratories".

But in some of those seven labs, the agreements have brought about major changes in the research culture. Topping the list is the National Chemical Laboratory (NCL) in Pune. Last year, it earned about US\$5 million, or 40% of its budget, from contracts with more than a dozen foreign clients. "NCL has its hands full with foreign projects," says one senior CSIR official, who asked not to be named. "I would not be surprised if they find no time to tackle problems posed by Indian industry." But Swaminathan Sivaram, head of polymers at NCL, rejects that criticism. He says he would be happy to enter into contracts with Indian companies, if only the demand was there. "Indian software professionals are working for America and nobody complains about it," Sivaram adds. "Whypick on us?"

K. V. Raghavan, director of the Indian Institute of Chemical Technology (IICT) in Hyderabad, also defends the rise in contract research. Besides bringing in money for his institute, he says it has exposed IICT scientists to the latest pharmaceutical

production methods. Raghavan argues that about half of the contracts the IICT has accepted will have relevance to India at "some time in the future".

But Raghavan's predecessor, A. V. Rama Rao, is less enthusiastic about the benefits of contract research to the institute. "No foreign company gives us any part of their core project," he says. Rama Rao now runs his own contract research company in Hyderabad, but argues that doing this work privately is different from committing the resources of publicly funded labs, which were set up using taxpayers' money with the intention of building up Indian industry's capabilities.

Balancing act

A single contract can employ a large number of researchers. The IICT, for example, deployed eight PhDs, 12 MScs and several technicians for one year on a SmithKline Beecham project for a total fee of US\$100,000. In the United States or Europe, this sum might not even cover the salary and lab costs of a single scientist. Even some senior figures at the IICT believe that directing high concentrations of human resources into contract research is damaging the institute's intellectual climate. "Pressure to earn money has arrested free thinking," says J. S. Yadav, deputy director of the IICT and head of its natural products division, and who manages many of the institute's research contracts. "Everyone is sitting with a calculator to see how much he will earn once the contract is done." Yadav also worries about PhD students who spend a year on a project that does not contribute towards their doctoral thesis.

Sensing something of a backlash against

his policies, Mashelkar accepts that contracts from foreign companies must not be allowed to dominate to the point that research institutes turn down offers of collaboration from Indian industry. "That day is very far off," he says. "But in any case we are already slowing down on contract research business."

As the CSIR tempers its enthusiasm for research contracts with western industry, the future of the labs set up by multinational companies on Indian research campuses is uncertain. Monsanto has already closed down the bioinformatics component of its Bangalore research centre. And Prem Vrat, acting director of the IIT, told *Nature* that companies must fall in line with a new 'synergy' policy or quit the campus when their rental agreements expire. The policy calls for openness, joint projects, two-way academic interactions and an annual review by a coordinating committee. "We will look at the conditions in the new policy and then decide our future course of action," says Pradeep Dubey, a spokesman for the IBM lab, adding that the company has yet to receive any formal communication from the IIT. But he admits that so far there has been no joint project with the IIT of the type that was originally envisioned.

Terminal problems

In part, the hostility towards IBM might reflect the financial anomalies its presence on the IIT campus has created. A student working for IBM in the summer can make more money than a professor working at the institute makes in a whole year. Not surprisingly, the lure of working for IBM has made some of the brightest students switch their priorities away from their supervisors' own projects. "When our students are sitting before the terminals, we do not know if they are working on IBM projects or those assigned by us," complains one computer scientist. In an attempt to ease these tensions, the IIT has now put a ceiling of five hours per week on the time students can work for IBM during term-time.

Given the benefits that collaboration with companies such as IBM could bring — not just money, but the chance to work on projects that otherwise might be beyond the reach of Indian research institutes — many of the country's academics hope that India's institutes will, in the future, manage to work with multinational companies on more favourable terms.

But there might be another way for Indian research institutes to profit from links between their brightest young students and western high-tech companies. The IIT is now eyeing some of its alumni who have made huge fortunes working in Silicon Valley. Perhaps donations from such individuals could provide an alternative to contracts from western companies. ■

K. S. Jayaraman writes for *Nature* from New Delhi.

Careers in science offer women an unusual bonus: immortality

Sir — I was alarmed to learn in your Opinion article¹ that President Clinton's National Science and Technology Council was "toothless" in its failure to address the shortage of women and minorities in science, technology and engineering, and that this situation could have "devastating" consequences by 2050 for the US economy and scientific leadership².

An analysis of death notices and obituaries in *Nature* every 10 years from 1949 to 1999, and in *Science* every 10 years from 1949 to 1969 (after which it stopped regularly publishing these) suggests a way of increasing the number of women scientists dramatically. As I show here, women scientists rarely die. Once word of this acquired immortality gets out, women should flock to scientific careers.

Of 1,184 obituaries in a three-year period coded for year of publication, sex, age at death, cause of death (if known) and field³, women accounted for 49 of 917 (5.3%) in *Science* and 13 of 267 (4.9%) in *Nature*; of the 44 commemorated in both journals, two were women. *Science* carried 3.43 times more obituaries than *Nature*; but the proportion of women remained constant at about 5% in each journal.

The dramatic increase in the number of women entering science, technology and engineering during the past 40 years (in which the number of female doctorates has grown at more than twice the rate of that for men, averaging 7.5% per year³) coincided with acquisition of immortality in increasing numbers of these women.

Although women in the physical sciences were represented by 4.8% of the death notices in *Science* and 8.3% of the obituaries in *Nature* in 1969, by 1979 there were none — they had become immortal (see Fig. 1). Since women received only 2.2% of US doctorates in engineering by 1978, more time is needed to assess the degree, if any, to which women in this field have acquired immortality. Women in the life sciences started to become immortal in 1979, but immortality is not yet fixed in this group, since one obituary appeared in 1999 — a year after women received 45.4% of the doctorates in that field (see Fig. 1). This trend is also found in other scientific and science-related fields of endeavour.

The fact that women were featured in some obituaries between 1949 and 1969 for all fields except engineering demonstrates that noteworthy women were contributing to scientific and scholarly endeavours half a century ago. As more females received doctorates over

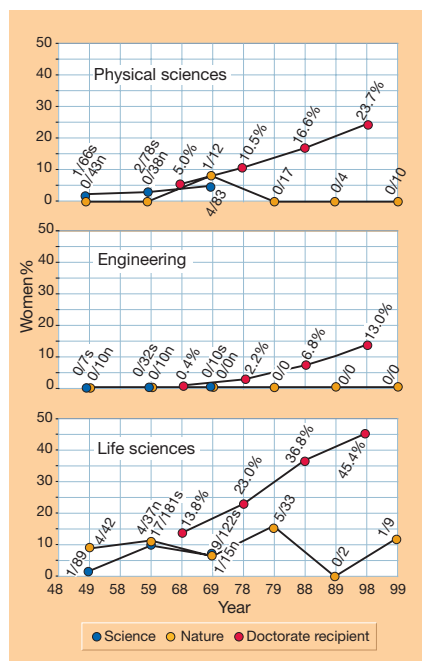


Figure 1 Percentages of women who received doctorates compared with those who received obituaries. Red circles, percentages of US doctoral degrees awarded to women during 1968, 1978, 1988 and 1998; blue circles, percentage of death notices for women in *Science* for 1949, 1959 and 1969; yellow circles, percentages of obituaries for women in *Nature* for 1949, 1959, 1969, 1979, 1989 and 1999. The numerator of fractions provides the number of obituaries for women; the denominator represents the total number of obituaries; s, *Science*; n, *Nature*.

subsequent years, however, the numbers of obituaries for women decreased to zero in the physical sciences, social sciences, education, humanities and other categories. One may therefore conclude that women in these fields no longer die.

The big question, of course, is what are the factors that led to their immortality? Is there a gene that predisposes women scientists to live for ever? If so, I propose the name *foy* (fountain of youth), and suggest that the researchers at DREADCO look into this.

Dean Falk

Department of Anthropology, University at Albany,
Albany, New York 12222, USA

1. *Nature* 404, 795 (2000).

2. Wadman, M. *Nature* 404, 800 (2000).

3. <http://www.nsf.gov/sbe/srs/srs00410/htmstart.htm>

Enigma thief stole a very special machine

Sir — Natasha Loder writes (*Nature* 407, 278; 2000) that the particular kind of four-rotor Enigma machine used by the German Abwehr — and stolen from

Bletchley Park earlier this year — is very rare, and that the only other known example is owned by the US National Security Agency.

This is correct. But, rare though the Abwehr versions are, there are several other four-rotor Enigma machines in existence and available to collectors by legal means. I myself own one: No. 877, bought at Sotheby's in March 1994 (it was previously sold at Phillips in April 1993). At least three others were sold at Phillips and Sotheby's during the 1990s.

E. T. Hall

Beenhams, Railway Lane, Littlemore,
Oxford OX4 4PY, UK

Did civil reactors supply plutonium for weapons?

Sir — We welcome the recent publication by the UK Ministry of Defence (MOD) of the first official inventory of the country's military plutonium^{1,2}. The report contains a remarkable admission²: "These figures show that the weapon cycle stockpile is in fact some 0.3 tonnes larger than the amount of plutonium the records indicate as available". Hence, the MOD was not aware of the existence of 60 bombs' worth of weapons-grade plutonium. The report does not attempt to identify the origin of this plutonium, simply quoting¹ "From unidentified sites, 0.37 tonnes", despite there being very few sources of weapons-grade plutonium.

We believe some calculations we published 15 years ago³ can help the MOD identify the source. In their early years (1963–72) the UK's civil Magnox reactors produced significant amounts of weapons-grade plutonium. In 1984 it was admitted that it was reprocessed at Sellafield in the same line, and at the same time, as the weapons-grade plutonium from military reactors⁴. British Nuclear Fuels Ltd, the plant operators, admitted that they called the weapons-grade plutonium "military" irrespective of origin⁴. It would have been consistent with these practices if all weapons-grade plutonium was shipped to the MOD's Aldermaston site. The government stated in 1983 that there was no weapons-grade plutonium in the civil stockpile⁵.

Today, the UK government refuses to quantify plutonium production from civil reactors for these early years. In 1985 we published an estimate of (0.36 ± 0.11) tonnes for the total weapons-grade plutonium produced by the UK civil reactors⁶. This agrees remarkably well with the MOD figure of 0.37 tonnes for plutonium of unknown origin. We conclude that about 11% of the

plutonium in UK nuclear weapons originated in civil reactors.

The MOD reports do not separate the transfer data into weapons-grade and non-weapons-grade plutonium, and there are no data on production in the country's dedicated military reactors at Calder Hall and Chapel Cross.

We call on the MOD to provide this information. Similar data have been made public in the United States³. The UK government is now in an anomalous position, having published the military stockpile while refusing to publish similar figures for civil plutonium. We request that they do so, and clarify the contradictory statements that have been made to Parliament about the fate of civil plutonium.

The Magnox reactors have entered their shutdown phase and are again producing significant amounts of weapons-grade plutonium. The UK government has recently decided to restrict information on plutonium production in civil reactors⁷. One hopes that history will not repeat itself.

K. W. J. Barnham*, **J. Nelson***,
R. A. Stevens†

*Physics Department, Imperial College of Science, Technology and Medicine, London SW7 2BW, UK

†Join Systems, Menlo Park, California 94025, USA

1. *Plutonium and Aldermaston: An Historical Account* (Ministry of Defence, London, 2000). <http://www.fas.org/news/uk/000414-uk2.htm>
2. *Historical Accounting and Plutonium* (Ministry of Defence, London, 2000). <http://www.fas.org/news/uk/000414-uk3.htm>
3. Barnham, K. W. J., Hart, D., Nelson, J. & Stevens, R. A. *Nature* **317**, 213–217 (1985).
4. Layfield, F. *Sizewell B Public Inquiry: Summary of Conclusions and Recommendations* (Department of Energy, London, 1987).
5. *Hansard* 27–7–83, cols 439–440 (1983).
6. *Plutonium: The First 50 Years* (Department of Energy, Washington DC, 1996).
7. Barnham, K. W. J., Nelson, J. & Stevens, R. A. *Nature* **395**, 739 (1998).

Achievers should stay to aid Brazilian science ...

Sir — The Opinion article “Genome sequencing for all” (*Nature* **406**, 109; 2000) exposed a patronising view of research in developing countries.

In my view, *Nature* could have used its valuable space to tackle more interesting, painful yet real issues surrounding scientists in developing countries (see the News feature “A springboard to success” in *Nature* **407**, 440–441; 2000). For example, why was the Brazilian paper celebrated in your Opinion article an exception rather than the rule?

Local antinationalism has allowed imperialism from industrialized countries to survive for centuries. So, although I understand the views of the Brazilian

scientists abroad “who frequently decide not to return, citing a lack of scientific opportunity”, they are also being used as cheap labour in rich countries. Hence they are perpetuating an unfair situation by their short-sightedness and selfishness (very often their studies have been funded by Brazilian public money).

By leaving Brazil they may well avoid having to carry out less ‘important’ or ‘glamorous’ science. But they also lose the chance to involve themselves in relevant issues such as the dismantling of Brazil’s public university system, or to claim the right to better jobs and working conditions, or to build a better future for themselves and for future generations.

Maria J. Hötzel

Universidade Federal de Santa Catarina, CCA-Departamento de Zootecnia e Desenvolvimento Rural, Rodovia Admar Gonzaga, 1346 Itacorubi 88.034.001, Florianópolis, Santa Catarina, Brazil

...yet the path is strewn with needless obstacles

Sir — I am a young Brazilian scientist and I agree with Tomas Prolla’s point in Correspondence¹ that the rigid bureaucracy in Brazil turns scientific research into a nightmare.

Four years ago, I asked the director of the institute where I did my PhD to release funds (about US\$20) so I could send reprints to England, as one of my results was going to be cited in a textbook to be published there. The director punched his desk and said that he was not there to support my megalomania. I sent the reprints using my own money, and my result was cited in the book². A professor from another university told me that to behave as I had done, at my level of seniority, would cause fear among my superiors.

I recently entered the selection process for a professor’s position in one of Brazil’s leading universities. One of the interviewers asked why I wanted to stay in academia instead of working in industry for better pay. I did not get the job.

The person who got the job has published about six papers in journals, and is corresponding author on none of these. My curriculum vitae lists 21 papers in good international journals. In 20 of these I am the corresponding author and in 11 I am the sole author. I have spent \$23,000 of my own money doing serious research in this country and I receive about \$200,000 as a government grant. In my laboratory I have the first atomic force microscope for biological research in the country.

This is the fourteenth selection process I have undergone in this country. Before I

received the result of my latest attempt, I was advised by another professor to go to the United States as I do not fit in the Brazilian system.

It seems that Brazil can produce good scientists for export, but this material does not bring income into the country.

Ricardo de Souza Pereira

Departamento de Parasitologia, Instituto de Ciencias Biomedicas, Universidade de São Paulo, Avenida Lineu Prestes 1374, Cidade Universitaria, São Paulo, Brazil

1. *Nature* **406**, 826 (2000).
2. Walker, G. *Yeast Physiology and Biotechnology* 16–17 (Wiley, Chichester, 1998).

If free speech costs lives that’s a high price to pay

Sir — Stewart *et al.* are right to remind us that the 17 years following the discovery of HIV have been a long time (*Nature* **407**, 286; 2000).

Both of us lost grandparents and great-grandparents to tuberculosis. One might have thought that in the hundred or more years since Robert Koch discovered *Mycobacterium tuberculosis* (without fulfilling all his postulates), we would have done a little better than the state we are in today: some 1.7 billion infected, with an annual death rate of 1.8 million. Of course we could all agree on tuberculosis being caused by another, as yet undiscovered, microbe riding on the intimate coat-tails of *M. tuberculosis*. Then perhaps the lack of progress would make sense. Paradigm lost.

In an earlier life one of us was valet to the French philosopher Voltaire. I remember cleaning his room one day, coming across a letter to Jean-Jacques Rousseau. As a Huguenot, I rejoiced at the remark, “I disapprove of what you say, but I will defend to the death your right to say it”. What is not widely known is the next sentence: “My only question, Sir, is whether the columns of *Nature* are appropriate?”

We are staunch believers in the right to free speech, but is *Nature* the appropriate place to militate in favour of the pre-Copernican model of the universe or the existence of phlogiston? After all, there is Speakers’ Corner in Hyde Park, when it’s not raining. To demand the right of reply or equal time on such matters is a trick the creationists have used.

HIV causes AIDS. Problems arise when the proposed alternative costs lives.

Simon Wain-Hobson*, **Robin A. Weiss†**

*Unité de Rétrovirologie Moléculaire, Institut Pasteur, 28 rue du Dr Roux, 75724 Paris cedex 15, France

†Windeyer Institute of Medical Sciences, University College London, 46 Cleveland Street, London W1T 4JF, UK

A many-feathered tale

Luck, and chickens, were pivotal in the identification of a dietary disease.

Beriberi, White Rice, and Vitamin B: A Disease, a Cause, and a Cure

by Kenneth J. Carpenter

University of California Press: 2000. 296 pp.

\$40, £24.95

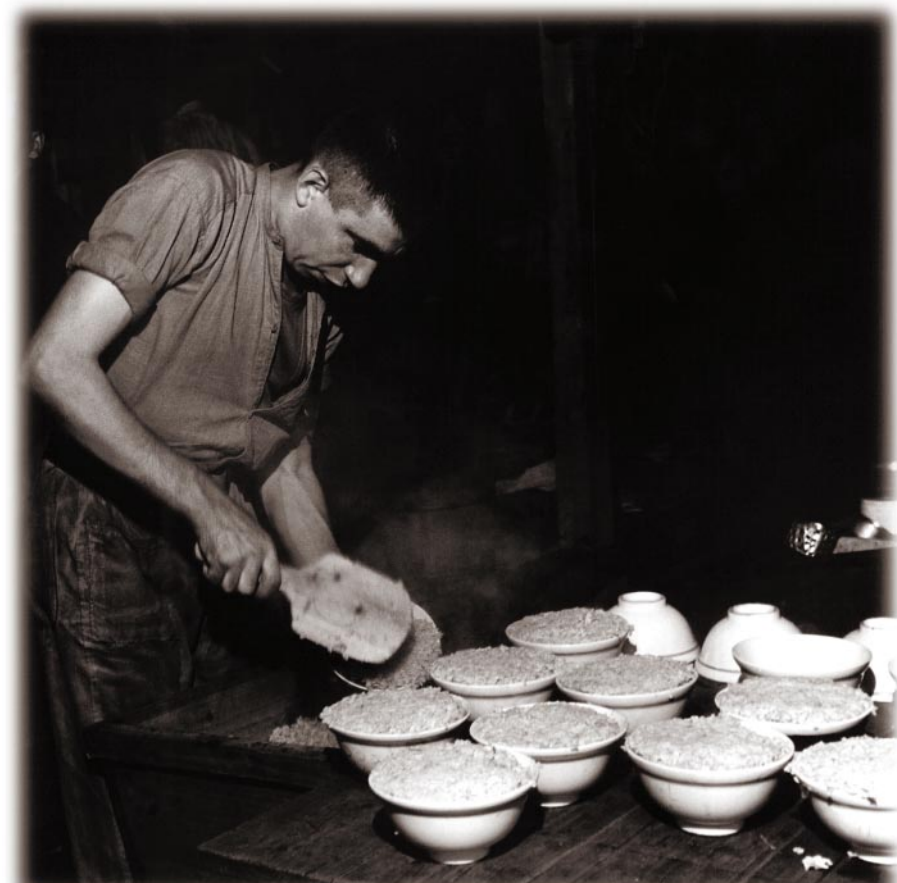
Susan E. Lederer

Beriberi is a disease that begins with weakness and loss of feeling in the legs, proceeds to swelling of the limbs and abdomen, and frequently ends in heart failure or death. At the biochemical level, the disease is caused by a deficiency of thiamine, or vitamin B. Once the scourge of millions of children and adults in Southeast Asia, thiamine deficiency is today recognized as one of the causes of mental degeneration in chronic alcoholics, the product of both a limited diet and impaired nutrient absorption. In a comprehensive and clearly written book, Kenneth Carpenter examines the origins of this 'tropical' disease and the detective work conducted on several continents that culminated in a cure and a preventive for beriberi.

The author of a book on the history of scurvy and vitamin C, Carpenter seeks to challenge the myth-making that has obscured the discovery of the cause of beriberi. Rather than telling the story of the lone 'scientist-hero' who, in a flash of inspiration, grasps the nature of the disease, Carpenter describes a considerably more crowded and interesting train of events.

In traditional accounts, the hero of the beriberi story is Christiaan Eijkman, a Dutch physician assigned to Cornelis Pekelharing's beriberi commission in Java in the late 1880s. In accordance with the reigning bacteriological paradigm, Eijkman attempted to infect chickens with blood taken from patients with beriberi. However, he noticed that uninjected chickens, fed on a diet of cooked white rice, developed the same leg weakness characteristic of the disease. Eijkman, the story goes, realized that beriberi is a function of the diet and not the work of a microorganism.

But Carpenter does not seek to topple Eijkman's claim to fame (Eijkman received the Nobel prize in 1929 for his work on beriberi). Rather, he provides a carefully nuanced account of the Dutch physician's experiments and their aftermath. Luck, Carpenter argues, played no small role in the chicken studies. Eijkman had begun his researches using rabbits and monkeys, but, needing large numbers of animals at low cost, he turned to chickens. A diet of cooked white rice would not have produced the same symptoms in rabbits that thiamine deficiency produced in fowl.



Jailbird: prisoners fed a white rice diet were, like the chickens, used to investigate beriberi.

Carpenter carefully analyses the extensive experiments Eijkman conducted on chickens, and notes, among other things, that after producing the leg condition in chickens fed on a white rice diet, Eijkman returned to Amsterdam, where he was not able to replicate his findings. At that point, he reverted to his earlier belief that the disease was caused by an infection, although by 1912 he was again convinced that the dietary explanation was correct.

Experiments involving animals and the search for an animal model of the disease form only part of the beriberi story. Carpenter examines the clinical trials conducted in prisons, hospitals and the military. Most of these human trials, he explains, were therapeutic in nature, efforts to alter the diet in some way to improve the health of individuals. In only one of the trials Carpenter considers did investigators deliberately assign prisoners to a diet expected to produce beriberi. In a Philippine prison in 1912, Harvard researchers Richard P. Strong and B. C. Crowell used 29 prisoners under sentence of death. Although the prisoners

received no reduction in their sentence, they did get unlimited free cigarettes and cigars for the period of the study. Carpenter does not mention that these 'volunteers' also signed agreements acknowledging that they risked developing beriberi as a result.

In a concluding chapter, Carpenter looks at the ethical questions in human and animal experimentation, describing this aspect as one of the main themes of the book. In most cases, however, the ethical issues raised by producing disease in large numbers of animals or the use of 'human guinea-pigs' receive only perfunctory attention.

But a major strength of the book is Carpenter's international perspective on the complex investigations into beriberi. In addition to the Dutch experimenters, he traces the contemporaneous and often competing teams of researchers in the United States, Britain, France and Japan, whose efforts contributed to solving the puzzle of the disease's aetiology, and to finding the means to treat the disease through vitamin supplements and enrichment of foods, and through social changes in the milling and preparation of rice.

Public-health policy regarding beriberi, as Carpenter reminds readers, is more than a matter of historical interest. In the 1980s Australian health policy-makers debated the utility of 'enriching' beer and other alcoholic beverages with thiamine as a means of increasing the thiamine intake of chronic alcoholics. Amid resistance from both the brewing industry and temperance organizations (who feared beer would be advertised as a health food), these plans were scrapped in favour of augmenting flour and bread. Here and throughout the book, Carpenter weaves the social, scientific and biochemical aspects of beriberi into a compelling story. ■

Susan E. Lederer is in the Section of the History of Medicine, Yale University School of Medicine, New Haven, Connecticut 06520, USA.

Cultivating the power of the empire

Nature's Government: Science, Imperial Britain, and the 'Improvement' of the World

by Richard Drayton

Yale University Press: 2000. 400 pp. \$40, £25

Lewis Pyenson

Today, English stands uncontested as the world's premier scientific language. This hegemony derives in part from the establishment of English-speaking institutions around the world at the time of the Industrial Revolution. Many of these institutions were concerned with exploiting nature for financial profit. In *Nature's Government*, Richard Drayton reviews how, beginning in the mid-eighteenth century, England established a large network of agricultural-improvement stations throughout its empire.

These institutions, orbiting around Kew Gardens, near London, centralized the botanical knowledge of exotic species, established protocols for moving cultivars around the world, and provided employment for hundreds of technical adepts, allowing many of them to write works on botany. Much of the money came from the public purse.

Unlike the omnibus research university, with its uncanny ability to capitalize on innovation wherever it appears, independent institutions with a particular research focus — from physic gardens and astronomical observatories to model farms and industrial testing laboratories — are perpetually threatened with superannuation. Government establishments in England that extorted profit from the soil of distant lands seem generally to have lagged behind the research front. Chemistry and then microbiology provided great promise for agriculture in the late nineteenth century; but their effects impinged on commercial botany only slowly and unevenly, as did Mendelian genetics in the twentieth century. At Kew, even biogeography and ecology seem to have found little place. *Nature's Government* is consequently heavy on government and light on nature. In the eighteenth century, Kew's founder Joseph Banks projected the gardens as an imperial repository for plant acculturation; his towering legacy proved hard to shake off.

Focusing on Kew, Drayton takes us from Banks, through the administration of the Hookers (William and Joseph) to the reign of William Thistleton-Dyer. The reader will discover how agricultural improvement and botany were part of a general move towards the professionalization of science in the nineteenth century. And Drayton's awkward prose informs the familiar story with new material.

It would have been good to know how all

the political rhetoric was reflected in botanical publications of a putatively dispassionate kind. Did discussion about commercial value find an echo in taxonomic choices and pathogenic identification? Or did botany, secure in its disciplinary norms, master its imperial patron? The answer would tell us much about the extent to which descriptive sciences and practical technologies are embedded in the social fabric of their time and place. ■

Lewis Pyenson is dean of the Graduate School, University of Louisiana at Lafayette, PO Box 44610, Lafayette, Louisiana 70504-4610, USA.

A warm topic for cogitation

The Hot Brain: Survival, Temperature, and the Human Body

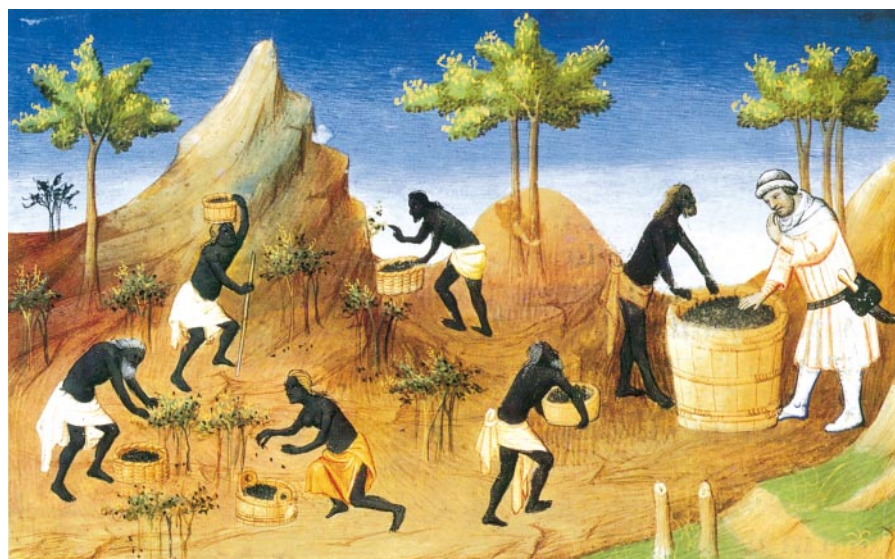
by Carl V. Gisolfi & Francisco Mora
MIT Press: 2000. 272 pp. \$45, £30.95

Andrew R. Cossins

Mammals and birds have much larger brains than do other vertebrates. They are also different in having warm and constant body temperatures. Is this a coincidence, or is there some formal or necessary link between the two facts? Perhaps the evolutionary development of a large brain required the attainment of homeothermy? Or could it be that the brain increased in size to serve the needs of homeothermy and only subsequently became invested with increased cognitive ability? Is the further increase in brain size during hominid evolution linked to coping with heat stress? These questions are given a detailed airing in an entertaining and provocative book that explores the evolutionary and functional link between brain, warmth and thermoregulation, with an emphasis on humans.

The evolutionary development from the thermotaxic responses of basal living forms to the complex homeostatic physiology found in mammals is given detailed treatment. And there is also extensive coverage of the way in which the hypothalamic and pre-optic regions of the vertebrate brain have evolved to become the main thermoregulatory control centres.

The issue of brain function and temperature affects a surprisingly large number of current issues in medicine and biology, and these liven up the book. For example, the acclimatization of humans to cold is explained within the context of the *ama* women divers of Korea and Japan who, before the advent of protective suits, acclimatized their bodies to the cold winter waters in which they dived. Evidence from studies in lower vertebrates suggesting that fever is adaptive and beneficial leads to the



Pioneering plunder: "The Harvest of Peppers in India" from Marco Polo's *Book of Wonders*.

question of whether it should be clinically relieved with aspirin. And does the loss of thermoregulatory control during REM sleep offer any insights into how human brains can be cooled without damage?

The book is entertainingly written, and touches on many issues of general interest. It would certainly capture the attention of biologists and medical practitioners, although, given the detailed accounts of neuronal structures, I doubt that it would be accessible to non-life scientists, despite the authors' intentions. Issues in human neuroscience are authoritatively covered, but the more comparative discussions of thermal biology and the evolution of homeothermy in mammals are less convincing. For example, the idea that vertebrate brains must be warm in order to function effectively does not accord with fish that live in cold, even polar, waters. Indeed, fish can display quite sophisticated thermoregulatory behaviour, as was shown more than 30 years ago, when goldfish trained to control water heaters by lever-pressing were able to regulate the temperature to within 1 °C. ■

Andrew R. Cossins is at the School of Biological Sciences, University of Liverpool, Derby Building, Liverpool L69 3BX, UK.

Catch as catch can?

Fish, Markets, and Fishermen: The Economics of Overfishing

by Suzanne Iudicello, Michael Weber & Robert Wieland

Earthscan/Island Press: 1999. 192 pp. £35/\$40 (hbk); £14.95/\$19.95 (pbk)

John Godfrey

Fish in the open sea are a resource common to all, but it does not pay any individual fisherman to hold back from overexploiting it. Indeed, if fish are becoming scarce, then fishing harder will often be the optimum short-term strategy. It is rational to overfish; this is 'the tragedy of the commons'. What is needed, if we are to avoid the tragedy, is common management, fairly enforced, so that the maximum yield that each population of fish can sustain indefinitely is never exceeded.

How to do this is far from obvious. Fishing policies of governments around the world have responded to pressure from fishermen, often to the detriment of the longer term. Overfishing brings short supply and high prices, to the disadvantage of the poor, for whom fish has historically been a source of cheap first-class protein. But consumers have little influence on policies, because their interest in conservation lacks the immediate impact that policies have on fishermen, and because they lack detailed knowledge. Their



Views of an underworld

The three oceans that surround New Zealand are home to a vast array of exotic marine life, including the Yaldwyn's triplefin (*Notoclinops yaldwyni*) show above, resting on pink sponge.

Underwater Wonders of New Zealand by Gillian and Darryl Torckler (New Holland, £19.99, \$32.95) is a lavishly illustrated guide to this underworld.

need, in both the short and long term, is sustainable management. The catching industry shares the same goal, but only in the long term.

Iudicello *et al.* bring a breath of fresh air to the often-stale fish dispute. They show how the exploitation of a new fish stock goes through clear phases: as the effort put into catching the fish increases steadily, the value of the catch grows quickly until the maximum economic yield is reached. There is then the greatest difference between the cost of fishing and the value of the catch. After this, profits fall off until fishing reduces the capacity of the stock to replenish itself. Beyond this, the population will fall and may crash. Policies for fishing should keep fish populations between the maximum economic and the maximum biological yield that can be sustained.

The book goes on to review examples of different management policies for various fish stocks around the world. Most have failed, often because governments give subsidies when fishing communities suffer as fishing becomes uneconomic. This lures back fishermen who might have left a depleted and uneconomic fishery.

The authors shy away from clear recommendations. But their valuable case studies suggest that somehow making fishing private property could be the answer. New Zealand has given individual quotas for fish

catches that can then be traded. But the tendency then is for the quotas to be bought up by a few big enterprises. Fishing has historically been a small-scale industry, with the benefits of fishing dispersed among communities with few alternative economic possibilities. If individual quotas were to revert to a regional management body, this difficulty could be avoided and the public interest taken into account.

Because the most productive waters are often fairly near to shore, the universal adoption of 200-mile Exclusive Economic Zones made rational and legally enforceable management possible. The European Union has pooled its EEZs, and although its common fisheries policy ought to be the most successful policy for Europe, it is not. For instance, one year's better recruitment to the cod stocks in the North Sea triggered a larger total allowable catch. The young cohort should instead have been allowed to grow and breed. But the Norwegians, not being members of the European Union and therefore unhampered by the common fisheries policy, have achieved a remarkable recovery of their overexploited cod. Another aspect is the fact that the water in the North Sea has become warmer, and less favourable to breeding by cod, so that greater care in management is needed.

The main failures of management have

been because the member states in the Council of Ministers so frequently water down the scientific advice they are given. The Treaty of Union requires that the common fisheries policy be reviewed in 2002, and this offers a golden opportunity for a radically improved policy. The European Commission will be producing a discussion document later this year. *Fish, Markets, and Fishermen* should help to stimulate fresh thinking on how to conserve fish better, for fishermen and consumers alike.

The open ocean presents an even more difficult problem, which is not seriously addressed in the book. Deep-sea fish live in cold water with sparse food supplies, and so grow and breed slowly. Although there are substantial standing stocks, they are particularly susceptible to overexploitation. As the boats and gear for catching them have required larger investments, there is strong motivation to overfish. And there is no sign that the Law of the Sea Convention will act soon enough.

John Godfrey, a member of the European Commission's Advisory Committee on Fisheries, is at 41 Lawford Road, London NW5 2LG, UK.

More on fishing

Fishing Grounds: Defining a New Era for American Fisheries Management

by the H. John Heinz III Center for Science, Economics and the Environment
Island Press, \$27, £25 (pbk)

Science in culture

Inverted logic

Antoni Gaudí's structural skeletons for Catalan churches.

Martin Kemp

It is something of a cliché to say that great thinkers can take an idea and turn it on its head in the search for new perspectives. But few can have done so as literally as the astonishing Catalan architect Antoni Gaudí, who designed a series of remarkable secular and ecclesiastical buildings in Barcelona in the early years of the twentieth century.

In common with adventurous architects across Europe, Gaudí was striving to create a form of design that was genuinely new — called art nouveau in France and modernism in Catalonia — yet deeply rooted in the past. Typically, designer-artists looked to styles that were local and indigenous, in Gaudí's case the medieval architecture of Catalonia and Spain's striking Moorish heritage. The intention was not so much the kind of eclectic imitation common in the nineteenth century as the extraction of underlying principles of form and structure.

To achieve this deeper penetration, a number of European designers were returning to what they considered to be first principles in the geometrical engineering of organic structures. Gaudí was convinced that the clustered piers and soaring vaults of a Gothic cathedral were closer to nature's living forms than the rectilinear geometry of classical columns and cornices.

In this pan-European quest for nature's geometry, Gaudí was inspired to research the complex geometry of surfaces, such as circular and hyperbolic paraboloids, so that he could construct intersecting combinations of convex and concave shapes that were rigorously geometrical and yet seemed true to nature.

When he translated this geometry into structural practice, his most original act was to re-characterize the arch according to the tenets of 'natural engineering'. He took as his model the catenary curve, formed when a chain hangs in a loop from two points. He deduced that the forces of tension in the catenary curve must be resolved along the line of the curve itself. If inverted as a 'catenary arch', the reverse forces of compression should run through the masonry of the arch, without causing the severe outward thrusts that generally necessitated the massive piers or flanking buttresses in large Gothic buildings.

Extrapolating this upside-down design method to cope with complex ecclesiastical buildings, Gaudí plotted their spatial structure through an intricate cobweb of wires and looped strings hanging from points on the scaled ground-plan of the projected building. Small, weighted sacks suspended from the lowest points of the string loops simulated, in reverse, the weight of the intricately curved vaults that were to be suspended between the skeleton of ribs.

Gaudí's towering Expiatory Temple of the Sagrada Família was begun in 1883 and was still piously under construction 74 years after his death. The remarkable hanging models he made for the temple and for his Chapel of the Colònia Güell (1905–15) no longer survive. But a reconstruction of the model for the chapel can now be seen in the museum attached to the unfinished temple (see figure). On such frameworks, duly inverted, Gaudí worked his plastic variations on organic morphologies. His aspiration was that the building should appear to the worshipper as a force of nature itself, a spiritual distillation of God's perfect engineering and design, rather than a mere container made by human agency.

In Barcelona today, Gaudí's legacy and the desirability of finishing his Sagrada Família are disputed by politico-religious factions in a way that ironically mirrors the architect's own journey from artistic radical to religious conservative. From an international perspective, we can see how he was one of the most original and profound advocates of reforming architectural design in the light of the timeless example of nature.

Martin Kemp is in the Department of the History of Art, University of Oxford, 59 George Street, Oxford OX1 2BE, UK.

Martin Kemp will be giving a talk, "not science AND art", at the Victoria & Albert Museum, London, on 3 November (7 pm).

Left, vaults of the Sagrada Família. Far left, Gaudí's hanging model (reconstructed) for the Chapel of the Colònia Güell (Gaudí Museum at the Sagrada Família, Barcelona).



Prophets without honour?

Or honour without profits?

R. M. Simmons

Two contrasting views of “The Future of Man” appeared in the British media in 1959. One was in the BBC Reith Lectures, given by the eminent British immunologist Peter Medawar — widely regarded among scientists as one of the most brilliant minds of the twentieth century — whose work on immunological tolerance laid the foundation for transplantation surgery and won him the Nobel prize in 1960. The other was a compilation of essays by the French palaeontologist Pierre Teilhard de Chardin, who, as well as holding several important posts in French science, was also a Jesuit priest. Teilhard’s literary reputation had already been made in France by an earlier book, *The Phenomenon of Man*, also published in English for the first time in 1959.

Medawar’s lectures dealt with the future of the human race from what he called the viewpoint of a human biologist. They were a lucid and closely argued account of the state of research on human genetics, fitness, intelligence and so on, and the prospects for improvement. At the outset he excused himself for not pursuing his “first thought of attempting a grand prophetic statement about Man’s future” on the grounds that “I would be obliged to weary you with endless qualifications and reservations and disclaimers, or else to try to disguise the thinness of the reasoning by taking refuge in apocalyptic prose”.

Teilhard had no such inhibitions: his view was that humans stood at the focal point of evolution, that there exists a kind of mental equivalent to the biosphere, the noosphere, which will eventually reach a critical point, analogous to self-consciousness. Sir Julian Huxley wrote an appreciative introduction for *The Phenomenon of Man*, and Arnold Toynbee wrote in the national newspaper *The Observer*: “This is a great book. Its subject is the sum of things; nothing less than God and the Universe. Teilhard sees and presents the Universe in evolution, but at the same time a unity. His vision of unity meets a spiritual need of our time.”

Medawar begged to differ: in 1961 he launched an attack on *The Phenomenon of Man* — which by this time had become a semi-popular classic — in the journal *Mind*; an article subsequently anthologized and often quoted. He successfully demolished Teilhard’s arguments in 11 pages of awesome, sustained invective. Or did he? Curiously, on close reading there is little real critical substance. He complains of Teilhard’s



Spot the heretic: Medawar (left) took a dim view of the Jesuit Teilhard de Chardin’s biology.

style (“tipsy prose-poetry”), some technical shortcomings (“no grasp of the real weakness of modern evolutionary theory”), but the main substantive issue is Teilhard’s misappropriation of scientific arguments to promote a religious standpoint (“obscure pious rant”) and so duping a gullible public (“educated far beyond their capacity to undertake analytical thought”). We shall never know what Teilhard thought of Medawar, as Teilhard died in 1954.

Teilhard’s books were published posthumously: his religious superiors forbade him publishing his views on human evolution in his lifetime. He thus shared with Galileo the distinction, if that is the right word, of having his work suppressed by the Roman Catholic Church. Just because Galileo was right does not, of course, give everyone else whose work is proscribed the stamp of scientific rectitude, but Teilhard was doubly distinguished in his second martyrdom at the hands of scientific orthodoxy. But apart from its religious streak, Teilhard’s approach is not so different from that of the

modern field of evolutionary psychology, and he anticipated the explosive growth of mass communication. For a book written in the late 1930s, *The Phenomenon of Man* seems remarkably prescient.

Teilhard is not alone in being tried by the scientific establishment while experiencing popular success. A good deal of hostility has been directed at the concept of the biosphere as an intelligent organism — James Lovelock’s Gaia — and at astronomer Fred Hoyle’s ideas on the extraterrestrial origin of life. Both met with popular enthusiasm before the scientific establishment would admit that they were candidate hypotheses. The evolutionary biologist John Maynard Smith castigated Hoyle recently in this very journal (*Nature* **403**, 594–595; 2000).

Teilhard’s books must have far outsold Medawar’s Reith Lectures, and therein lies a dilemma for scientists in their relationship with the public. Should they, like Medawar, stick to the facts, satisfying the dictates of scientific conscience but, with a limited horizon, reaching a limited audience? Or should they throw caution to the winds as Teilhard did, appeal to a large audience, but risk disapprobation by the scientific community? There is a psychological issue, too, which is that the public may have twigged that not only do orthodox scientists restrict their enquiries to the physical world, but also that many of them believe in their hearts that there is nothing beyond it. ■

R. M. Simmons is at the Randall Centre, King’s College London, Guy’s Campus, London SE1 1UL, UK.

Teilhard was doubly distinguished in falling foul of both the Catholic Church and scientific orthodoxy.

Subpoenaed in Syracuse

Eureka!®

Tom Holt

“Da-da-da-dum,” sang Archimedes, reaching over his shoulder with the sponge to get at the awkward spot in the middle of his back. “Da —”

Careless — a moment’s thoughtlessness, and there’s a big pool of water on the floor. Damn. Pause. Why is there a big pool of water on the floor? The philosopher frowned. An insight, nebulous as a cloud, began to condense in his mind.

“Hold it!” said a voice from behind.

Archimedes jumped, thereby increasing the size of the pool and, coincidentally, proving the hypothesis he’d been working on. “Who the hell are you?” he asked.

The stranger stepped out of a fold of shimmering blue light. “You don’t know me,” he said. “My name’s Calvin Dieb. I’m a lawyer.”

Archimedes stared at the blue light, the stranger’s outlandish clothes. “Are you a god?” he asked.

“Nah. Easy mistake to make, though,” the stranger reassured him.

“Actually, I’m from the future. Three thousand, one hundred and fifty years, to be exact. In my century, we’ve figured out how to travel backwards in time. Oh, forget I said that, by the way.” He chuckled. “Don’t want that outstandingly intuitive mind of yours getting on the job back here in the Dark Ages, could lead to serious doo-doo. So,” he went on, stepping clear of the blue fire and vanishing it with a click of his fingers, “this is it, then. The big moment. Congratulations.”

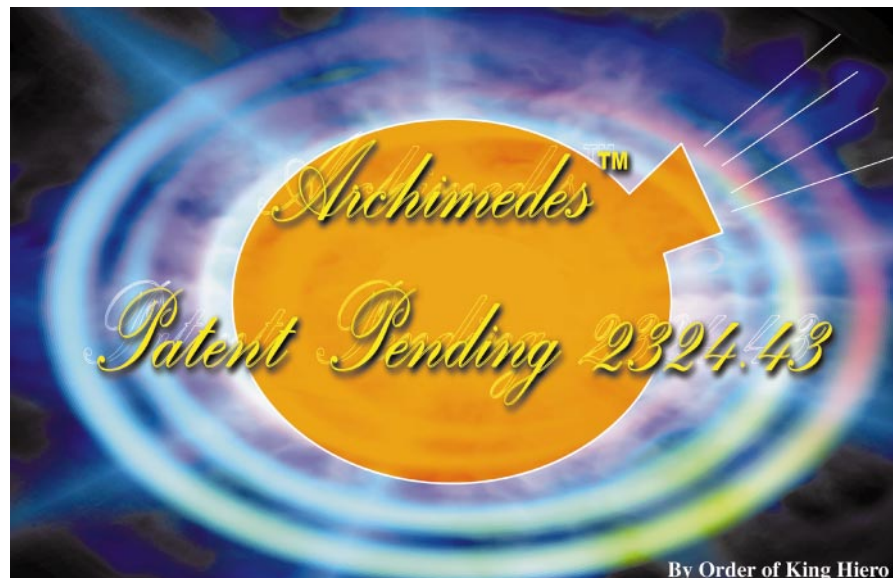
“Is it?”

“Sure.” The stranger grinned like an open wound. “Because of this, your name’s gonna be a household word for the rest of Time — believe me,” he added with a wink, “we checked. This discovery of yours, it’s gonna revolutionize the way mankind understands nature. It’s practically the birth of science. And you know what really burns me up about it?”

Archimedes thought for a moment. “Well, no,” he said.

“What really bugs me is,” said the stranger, “you don’t make so much as a wooden nickel out of the whole deal. Not a cent. Nada. One of the most seminal discoveries in human history, and the guy who made it has to go on washing his own tunics. Now I ask you, is that right?”

Archimedes thought some more. “Yes,” he said. “I mean, it’s interesting, I suppose, in a kind of bet-you-didn’t-know-that sort of



way, probably get me invited to a few parties, but it’s not as if it’s any good for anything —”

The stranger snorted. “That, my friend, that is where you’re — with the greatest respect — totally wrong. I can’t go into details for fear of screwing up the timelines, but trust me, this is gonna be huge. Multi-billion-drachma huge. And who’s gonna get all that money? Not you, friend. Not,” he added, leaning forward a little, “unless you listen to what I’ve got to say.”

Archimedes frowned. “I’m listening,” he said.

The stranger nodded. “Back where I — forward when I come from,” he said, “we got a thing called patents. Means that if someone wants to use your idea, they gotta pay you money.”

“Really? What a strange idea.”

“It’s cool,” the stranger said enthusiastically. “Now, what I’m proposing is, in return for a small piece of the action, say one third, I explain to you how it works. In the morning you go and see your friend King Hiero and explain to him how, if he passes a law here in Syracuse whereby inventors like yourself get paid a whole lot of money each time they invent something, pretty soon Syracuse’ll be the technological and economic capital of the world, and he’ll be the biggest king, bigger than Rome and Macedon and Carthage put together. Then we patent this new discovery of yours, and after that it’ll just be a matter of raking in the dough.”

After the stranger had finished explaining the theory and practice of patent law, the philosopher’s eyes were burning like stars. “That’s brilliant!” he exclaimed. “I’ll go and

see the king right this minute.” And he jumped out of the bath and headed for the door, pools of displaced water forming unheeded at his every step.

“Hey,” the stranger called after him, “aren’t you gonna get dressed first?”

Archimedes came home six hours later, in the king’s personal chariot, laden with gifts, and immediately signed all six copies of the contract that the stranger shoved under his nose.

“Great,” the stranger said, as he clicked his fingers to light the blue flames of the temporal interface. “You won’t regret this.” He snapped his fingers again. Nothing happened.

Archimedes’ invention of the patent in 221 BC revolutionized the scientific world. Instead of blurting out discoveries for anybody to hear, philosophers revealed them only to the rich merchants who had the wealth to develop them properly. Because they were merchants rather than scientists, they chose to finance the projects that looked to them as if they promised a good, quick return on capital. Mankind never did discover gravity, but the whoopee cushion was invented in 146 BC.

As for Calvin Dieb, he made the best of a bad job, eventually settling down in the small city of Acragas where he tried to interest the local goat-farmers in product liability litigation. Ironically, he was killed by an inventor, striving to patent the secret of flight, who fell on him after flying too close to the Sun and melting the wax that held his wings together. ■

Tom Holt used to be a lawyer, but he’s better now. His latest novel, *Valhalla*, is published by Orbit.

Identifying cosmic muck

Harry Y. McSween Jr

A fireball that landed in the Canadian Yukon has yielded meteorite fragments that are proving difficult to classify — could they be the most primitive samples to have been found on Earth so far?

Meteorite falls are common enough, but only rarely is a new type of meteorite discovered. Last week in *Science*, Brown *et al.*¹ described what might be such an event with the orbit, fall and recovery of numerous fragments of a meteorite in the Canadian Yukon. Named after the frozen lake on which it fell, the Tagish Lake meteorite is a chondrite, which means it is a stony aggregate of matter from the early Solar System. It is described as a carbonaceous chondrite — one which contains relatively high levels of carbon — but it also has unusual properties that make it difficult to classify properly.

All chondrites are samples of matter from the early Solar System that have undergone only minor geological processing. As such they can provide us with a snapshot of some of the earliest material in the Solar System. But the chemical composition of the Tagish Lake meteorite suggests that it might be especially primitive. In other words, its composition is very similar to that of the Sun.

Much of what we know about the abundance of elements in the Sun is derived from analysis of carbonaceous chondrites, some of which bear more than a passing resemblance to a dried-out mud puddle. The luminous centrepiece of our Solar System would appear to have little in common with a lump of mud. But spectroscopic estimates of elemental abundances in the Sun's visible surface, known as the solar photosphere, match most closely with the compositions of these meteorites. Consequently, trace elements in the Sun that cannot be measured spectroscopically are determined using elemental abundances in chondrites^{2,3}.

Despite their primitive chemistry, carbonaceous chondrites typically contain clays, oxides and other minerals, sometimes cemented by veins of carbonate and sulphate⁴. Mineralogically, they are more 'cosmuck' than cosmic. The mineral alteration is thought to occur within the parent asteroids, when accreted ices are melted by heat produced during the decay of short-lived radioactive nuclides⁵. Like other carbonaceous chondrites, the Tagish Lake meteorite has reacted with water at low temperatures, but its mineralogy appears to be somewhat less altered (see Box 1).

Brown and colleagues¹ speculate that the Tagish Lake meteorite might be a sample of the material that was the precursor of

Box 1 Making the grade

As their name suggests, carbonaceous (type C) chondrites generally contain more carbon than the ordinary chondrites (such as H, L or LL), the most common stony meteorites. Chondrites are classified according to their chemical group (given by a letter) and their petrologic type (given by a number). So far, six chemical groups of carbonaceous chondrites have been identified: CI, CM, CR, CV, CO and CK (where the second letter refers to the name of a prominent member of each group). Each chemical group probably represents chondrite samples from a single asteroid body. The CI chondrites are the most primitive; in other words they are closest in composition to the Sun.

Petrologic type indicates the degree of mineralogical or textural modification that each sample has experienced through geological processes

	Petrologic type	Chemical group
Aqueous alteration ↑	1	CI
	2	CM, CR
	3	CV, CO, H, L or LL
Thermal metamorphism ↓	4	CK, H, L or LL
	5	H, L or LL
	6	H, L or LL

on its parent asteroid. Some chondrites have been heated and metamorphosed, producing new high-temperature minerals and recrystallized textures, whereas others have suffered aqueous alteration by fluids at low temperatures. In the figure, increasing thermal alteration is indicated by petrologic grades 4–6, whereas grades 2 and 1 indicate increasing degrees of low-temperature alteration of the meteorite by water. So type 3 chondrites are the least modified, type 1 chondrites are

the most altered by water, and type 6 chondrites the most metamorphosed by heat. All CI chondrites are highly altered (type 1), but they are chemically the most primitive group. Brown *et al.*¹ suggest that the Tagish Lake meteorite may be a new class of carbonaceous chondrite that acted as the precursor to CI1. If this meteorite turns out to be a CI2, as Brown *et al.* say it could be, it would truly be a less altered sample of the most primitive meteorite group. **H. Y. M.**

the CI chondrites, the group of chondrites until now thought to be the most chemically primitive group. As such, they say, it may be a new type of carbonaceous chondrite.

The relative proportions of oxygen isotopes vary considerably among chondrites and are commonly used in meteorite classification. The oxygen-isotope composition of Tagish Lake places it among the CI chondrites (see Box 1). But in terms of texture, the meteorite resembles another carbonaceous chondrite group, the CM chondrites. As a result, this meteorite may force meteoritists to revise their classification system.

It is the chemistry of the Tagish Lake meteorite, rather than its mineralogy, that is probably most intriguing. Samples of the still-frozen meteorite were collected soon after it fell. As a result, they have probably suffered minimal terrestrial contamination. In contrast, museum samples of many CI and CM chondrites have been stored under

less than sterile conditions. Brown and co-workers have analysed the abundances of most elements in the Tagish Lake material, and their analyses offer an interesting comparison with the Sun's composition. Meteoritists divide elements according to their volatility, because elements are partitioned among and within chondrites based on the temperatures at which they condensed. The abundances of the least volatile elements in Tagish Lake resemble those in CM chondrites, whereas the more highly volatile elements lie between CM and CI values.

Comparing elemental abundances, Brown and co-workers found that CI chondrites provide a closer match to solar composition for 39 analysed elements, whereas Tagish Lake provides a closer match for just 17 elements. From this comparison, it is not obvious whether Tagish Lake is more primitive than CI chondrites, as the most primitive sample would be expected to match the

Sun's composition most closely. But the arguments here may be somewhat circular because solar abundances are estimated partly from CI chondrites. Trace elements in the Sun that are determined from chondrites appear in both lists of 39 and 17 elements so they do not help us to determine which rocks are the most primitive. In any case, analytical uncertainties in elemental abundances for the solar photosphere are relevant to the composition of both the Tagish Lake and other CI chondrites.

I have always been intrigued that the best match for solar abundances is provided by heavily altered meteorites that are basically congealed mud puddles. It is perplexing that even a less altered carbonaceous chondrite such as Tagish Lake should preserve its primitive chemical composition. Many of the meteorite's former minerals have been replaced by hydrated phases, and abundant carbonates have precipitated from solutions percolating through the stone.

On Earth, mineralogical alteration by fluids wreaks havoc with rock chemistry, and

oxygen isotopes in the Tagish Lake material indicate that water and rock exchanged oxygen atoms freely. It is difficult to see how the isotopic composition of oxygen, which makes up almost half the rock by weight, could have been modified without affecting other elements in lesser proportions. Some might argue that in a closed system — such as on the parent asteroid — thermal and aqueous processes could alter the mineralogy of a rock without affecting its overall chemistry. But just how a soggy meteorite retains its cosmic composition is a question we still cannot answer. ■

Harry Y. McSween Jr is in the Department of Geological Sciences, University of Tennessee, Knoxville, Tennessee 37996-1410, USA.
e-mail: mcsween@utk.edu

1. Brown, P. G. *et al.* *Science* **407**, 320–325 (2000).
2. Anders, E. & Grevesse, N. *Geochim. Cosmochim. Acta* **53**, 197–214 (1989).
3. Burnett, D. S. *et al.* *Geochim. Cosmochim. Acta* **53**, 471–481 (1989).
4. Zolensky, M. E. & McSween, H. Y. Jr *Meteorites and the Early Solar System* 114–143 (Univ. Arizona Press, 1988).
5. Grimm, R. E. & McSween, H. Y. Jr *Icarus* **82**, 244–280 (1989).

There have been previous reports of seemingly ancient bacteria from environmental samples, such as ancient rocks and coal, and bricks from 2,400-year-old Egyptian temples². But these claims were met with scepticism: early research on bacterial survival in soil on herbarium samples from Kew Gardens, London, stored and added to regularly since 1640, suggested that even spores were likely to survive for only about 1,000 years³. In addition, since the 1960s there have been several reports of bacteria in ancient salt deposits, varying in age from 195 million to 650 million years (see refs 5, 6 for reviews). But there have been considerable doubts about the origin of these bacteria. One concern is the possibility of contamination by ubiquitous modern bacteria, during either sampling or subsequent handling. Another is that the bacteria may have entered the formations where they were found some time after the formation was deposited. There is also the question of how bacterial biopolymers can remain intact over millions of years in dormant bacteria⁷; or, conversely, if bacteria are metabolically active enough to repair biopolymers, this raises the question of what energy source could last over such a long period.

A breakthrough came in the mid-1990s with the detection of bacterial spores in a bee preserved in amber 25–40 million years old⁸. The authors isolated DNA from these spores, and it turned out to be similar to that of the *Bacillus* species they cultured from the same bee after the surface of the amber had been sterilized.

Vreeland and colleagues used a similar approach. They thoroughly sterilized the surface of their samples. They then drilled into and removed fluid from a tiny inclusion (about 9 microlitres in volume) and inoculated the fluid into growth medium. The bacterium they isolated was also from the genus *Bacillus*, and was also doubly protected — first within the fluid inclusion (Fig. 1b of the paper on page 897), rather than a bee, and then within the salt crystal, rather than amber. Their results support the view that ancient bacterial spores can survive for millions of years, but remarkably

Microbiology

A case of bacterial immortality?

R. John Parkes

On page 897 of this issue Vreeland *et al.*¹ describe the cultivation of a bacterium from the past. Nothing new about that, you might think. In adverse conditions, some microorganisms form resistant structures called spores that are known to survive for considerable periods of time.

For instance, spores have been found in a 118-year-old can of meat²; and a yeast has

been cultured from a 166-year-old bottle of porter ale, and was subsequently used to reproduce the beer³. Findings such as these are rather disconcerting, for we rely on sterilizing or inhibiting the growth of microorganisms in food to prevent food poisoning and product deterioration. Most microbial cultures, however, are killed quite easily by heating, chilling or freezing, disinfection or even — and this is pertinent to Vreeland and colleagues' paper — an abrupt change in salinity (hence the food-preservation techniques of canning, freezing, pickling and salting). But what is new is that Vreeland *et al.* claim to have isolated bacteria from 250-million-year-old, buried salt crystals. If bacteria can survive for this length of time, why should they die at all?

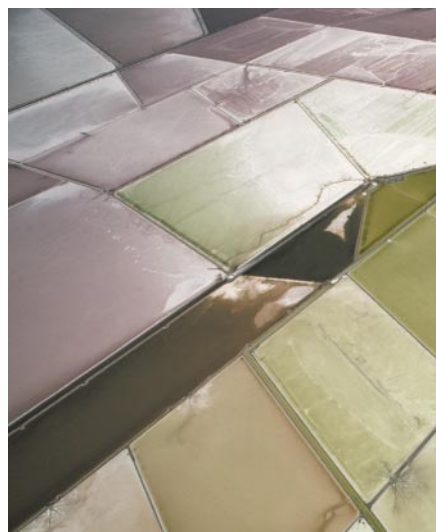


Figure 1 Evaporation pans in the Great Salt Lake, Utah, United States. The pink ones are where evaporating salt water has reached saturation and halophilic archaea are thriving.



Figure 2 Fluid inclusion in a 20-million-year-old salt crystal, in visible (left) and ultraviolet light. Fluorescence under ultraviolet is consistent with the presence of organic matter. But it is not conclusive evidence, nor does it provide information about the type of material — which might be the remains of entombed bacteria or, conceivably, the intact biopolymers of ancient but viable bacteria. The crystal is about 65 micrometres in length. (Image courtesy of W. D. Grant.)

extend the apparent timescale from tens to hundreds of millions of years.

The presence of spores in such old salt crystals is plausible, as spores are specialist survival structures produced by a few bacterial groups including bacilli⁹. They have a thick protective protein coat, and the cytoplasm is partially dehydrated and mineralized, causing enzymes to become inactive, and the DNA is stabilized.

Vreeland and colleagues' isolate (2-9-3) grew in an organic medium with 20% salt, and came from one of only three crystal inclusions (out of 66) that resulted in bacterial growth. So bacteria in inclusions are relatively rare. Genetically, isolate 2-9-3 is closest to another bacillus, *B. marismortui*, which was isolated from, and called after, the (clearly inappropriately named) Dead Sea¹⁰. Notably, however, *B. marismortui* was not isolated from freshly collected Dead Sea water, but from water sealed in a bottle 57 years ago. The organism is a moderate halophile that can grow in salt concentrations only up to 25%.

The growth conditions for the 2-9-3 isolate have yet to be determined, and the authors describe it simply as salt (halo-) tolerant. It will be fascinating to see if its halotolerance is similar to that of *B. marismortui*. If it is, then 2-9-3 could not have grown in an evaporating salt pan when concentrations approached saturation (about 30%) and salt deposits began to form. These worsening conditions would have triggered spore formation, and the spores could have become trapped in tiny pockets of brine, in the inclusions that are a permanent feature of salt crystals. It seems they were subsequently buried by further halite formation to a depth of 569 m, to be retrieved by Vreeland *et al.* and revived from their 'cryptobiotic' state some 250 million years later. Interestingly, the first report of a bacterium from an ancient salt formation was also of a spore-forming bacillus, *B. circulans*¹¹. This was proposed to be even older (650 million years old). But without the careful contamination checks of Vreeland *et al.*, the results were viewed with scepticism.

When evaporating sea water reaches saturation, moderately halophilic bacteria are replaced by representatives from the archaeal domain of life. These halophilic archaea thrive under such conditions, reaching concentrations of more than 10^8 cells per gram (Fig. 1). So it might not seem surprising that archaea have also been repeatedly cultured from ancient brines and occasionally from surface-sterilized crystals^{4,12} (but not yet from fluid inclusions, although they can be trapped within them¹³). However, halophilic archaea do not form protective spores, so how they could survive for millions of years is an even greater mystery. One suggestion⁵ is that the slow movement of fluid inclusions through rock salt could replenish organic

substrates and nutrients, and thus enable survival (Fig. 2).

Given the care that Vreeland *et al.* took to avoid contamination, their results are the best evidence yet for the extremely long-term survival of microorganisms. The potential implications are profound. For instance, can spores effectively be immortal? What is the biochemistry that allows them to survive for so long? Where else on Earth, and to what depths, might ancient bacterial life be lurking¹⁴? And, given this startling example of apparent bacterial durability, do spores in rocks even provide a mechanism for life to be transported between planets by 'panspermia', as has been proposed¹⁵?

Before the results of Vreeland *et al.* can be fully accepted, however, they will require replication by other groups and similar studies to be carried out on different salt deposits. In addition, the presence of spores within inclusions needs to be demonstrated by direct microscopy, and the DNA sequence of the 2-9-3 isolate will have to be detected in the fluid from the inclusion. But irrespective of their origin, halophilic bacteria can clearly be isolated from rock

salt. So the next time you sprinkle salt on your food, think what else you might be eating.

R. John Parkes is in the Department of Earth Sciences, Wills Memorial Building, University of Bristol, Bristol BS8 1RJ, UK.

e-mail: j.parkes@bristol.ac.uk

1. Vreeland, R. H., Rosenzweig, W. D. & Powers, D. W. *Nature* **407**, 897–900 (2000).
2. Postgate, J. *New Sci.* 21 July, 46–49 (1990).
3. Kennedy, M. J., Reader, S. L. & Swierczynski, L. M. *Microbiology* **140**, 2513–2529 (1994).
4. Sneath, P. H. A. *Nature* **195**, 643–646 (1962).
5. Grant, W. D., Gemmell, R. T. & McGenity, T. J. *Extremophiles* **2**, 279–287 (1998).
6. McGenity, T. J., Gemmell, R. T., Grant, W. D. & Stan-Lotter, H. *Environ. Microbiol.* **2**, 243–250 (2000).
7. Lindahl, T. *Nature* **362**, 709–715 (1993).
8. Cano, R. J. & Borucki, M. *Science* **268**, 1060–1064 (1995).
9. Madigan, M. T., Martinko, J. M. & Parker, J. *Brock Biology of Microorganisms* 8th edn (Prentice Hall, London, 1997).
10. Arahall, D., Marquez, M., Volcani, B., Schleifer, K. & Ventosa, A. *Int. J. Syst. Bacteriol.* **49**, 521–530 (1999).
11. Dombrowski, H. J. *Parasit. Infek. Hyg. Abt.* **178**, 89–90 (1960).
12. Norton, C. F., Grant, W. D. & McGenity, T. J. *J. Gen. Microbiol.* **139**, 1077–1081 (1993).
13. Norton, C. J. & Grant, W. D. *J. Gen. Microbiol.* **134**, 1365–1373 (1988).
14. Wellsbury, P. *et al.* *Nature* **388**, 573–576 (1997).
15. Davis, P. *The Fifth Miracle: The Search for the Origin of Life* (Penguin, London, 1998).

Cardiovascular biology

Tuning channels for blood pressure

Nick Standen

Information is transmitted within cells by chemical signals, which may act locally (near to where the signal is produced) or globally (throughout the cell). These signals act on molecular targets, which are tuned and positioned to receive the correct signal and respond to it properly. Calcium ions are particularly important cellular signals that can regulate functions ranging from gene expression to muscle contraction. Acting globally, calcium ions can induce the contraction of smooth muscle, causing an increase in blood pressure within arteries. Acting locally, calcium works through an ion channel (a protein that permits and controls ion flow across cell membranes) to cause the relaxation of smooth muscle. On page 870 of this issue¹, Brenner and colleagues reveal how the ion channel in question is tuned to respond to local calcium signals.

The blood pressure within large arteries is controlled by the resistance of small arteries and arterioles (even smaller arteries) to blood flow. Constriction of these blood vessels raises blood pressure in the same way that pinching the end of a garden hose raises the pressure within it. The diameter of the blood vessels depends on the contraction of smooth muscle cells in their walls, and this in turn is controlled by the concentration of Ca^{2+} ions within the cells.

Calcium has several roles in controlling contraction (Fig. 1, page 847). First, the global Ca^{2+} concentration determines contraction by acting through Ca^{2+} -dependent enzymes, which in turn affect the contractile proteins. The global Ca^{2+} concentration is determined in part by the rate at which Ca^{2+} enters the cell through channels in the plasma membrane (Fig. 1a). These channels are regulated by the voltage across the membrane; the voltage inside the cell is usually around -40 millivolts, and the channels open as this becomes more positive. So the voltage across the plasma membrane is a key determinant of contraction².

Second, Ca^{2+} can act locally (Fig. 1b). Ca^{2+} can be released from an intracellular compartment, the sarcoplasmic reticulum, through a Ca^{2+} -permeable channel on its membrane. Named for its sensitivity to a plant alkaloid, this channel is the 'ryanodine-sensitive Ca^{2+} -release channel'. It is actually activated by Ca^{2+} found outside the sarcoplasmic reticulum (Fig. 1b). Local Ca^{2+} signalling may be important here, as the gap between plasma membrane and the sarcoplasmic reticulum is only 20 nanometres. So it is possible (although not certain) that Ca^{2+} , entering the cell through the voltage-sensitive channels on the plasma membrane, activates the ryanodine-sensitive channels

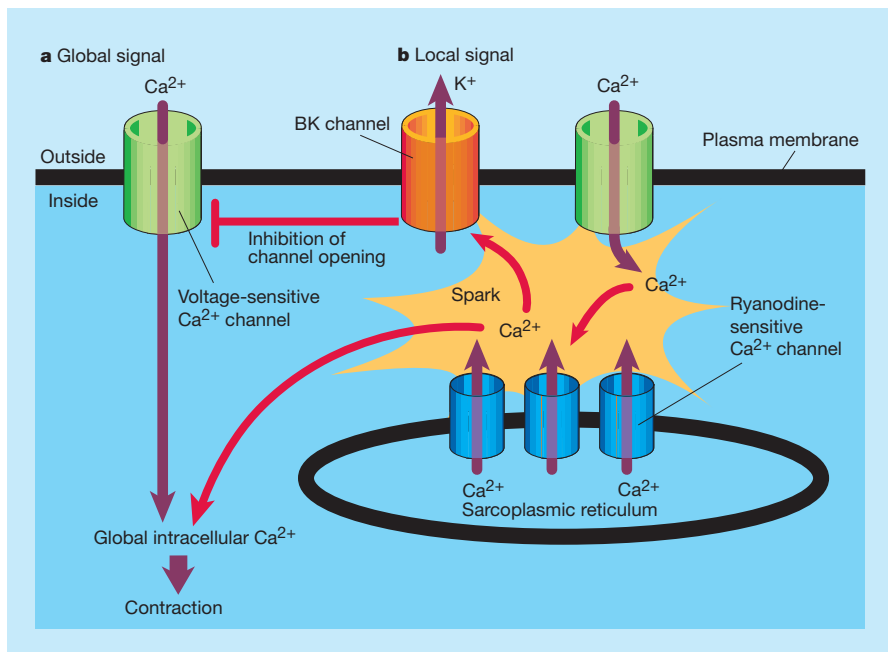


Figure 1 Global and local calcium signals control the contraction of smooth muscle and thus blood pressure within arteries. **a**, Global signals act throughout the cell. Ca^{2+} enters smooth muscle cells through voltage-sensitive Ca^{2+} channels on the plasma membrane, contributing to the global intracellular Ca^{2+} concentration which controls contractile force. **b**, Local signals act near to the site at which they are produced. Ca^{2+} can be released from the sarcoplasmic reticulum through ryanodine-sensitive Ca^{2+} -release channels, giving rise to a local Ca^{2+} 'spark'. These channels may be activated by Ca^{2+} entering through voltage-sensitive channels on the plasma membrane. 'Big K⁺-conductance' (BK) channels are activated by this spark to release K^{+} ions, making the voltage across the plasma membrane more negative. This inhibits the voltage-sensitive channels, reducing both the global Ca^{2+} concentration and contraction. Brenner *et al.*¹ have identified the $\beta 1$ subunit of the BK channel as being that which tunes the channel to the local Ca^{2+} spark.

on the sarcoplasmic reticulum by means of a sub-plasma-membrane rise in Ca^{2+} (ref. 3). In any event, concerted Ca^{2+} release from the sarcoplasmic reticulum through a group of ryanodine-sensitive channels produces a temporary local increase in Ca^{2+} concentration, called a ' Ca^{2+} spark'³. Such sparks can control contraction, but how do they do this?

The Ca^{2+} target studied by Brenner *et al.* is a calcium-activated, potassium-permeable channel in the plasma membrane. This channel is called the 'big K⁺ channel' (BK channel), because K^{+} ions flow through it at a high rate. BK channels in smooth muscle cells respond to local Ca^{2+} sparks rather than global Ca^{2+} signals⁴. The sparks open BK channels, and the resulting outward flow of K^{+} ions makes the interior voltage temporarily more negative. This then closes the voltage-dependent Ca^{2+} channels in the plasma membrane, providing a brake on contraction. So the global and local calcium signals have opposite effects on contraction: high global concentrations of Ca^{2+} activate contraction, whereas Ca^{2+} sparks inhibit it.

BK channels comprise four so-called α -subunits, surrounding the K^{+} pore, usually together with an unknown number of β -subunits — of which four different types

($\beta 1$ – $\beta 4$) are known so far. BK channels in smooth muscle cells are thought to contain the $\beta 1$ subunit⁵. In artificial BK channels, the $\beta 1$ subunit increases the channel's sensitivity to Ca^{2+} and slows the kinetics of opening and closing of the channel⁵.

Brenner *et al.*¹ now suggest that the $\beta 1$ subunit tunes the kinetics and the Ca^{2+} sensitivity of the BK channel so it opens appropriately during the high local Ca^{2+} concentrations achieved during a Ca^{2+} spark. The authors genetically engineered mice so they lacked the $\beta 1$ subunit, and found that this disabled the negative-feedback system induced by sparks. The Ca^{2+} sensitivity of the BK channels is reduced in the genetically engineered mice, and the channels are no longer tuned to Ca^{2+} sparks, which occur as normal but lead to little BK-channel activation. The physiological consequences are severe: cerebral arteries, and presumably others, are more constricted, so the arterial blood pressure is permanently high; the mice also develop enlarged hearts, a feature commonly seen in humans with hypertension (high blood pressure).

The regulation of BK channels by local rather than global Ca^{2+} signals requires that these channels be appropriately sensitive to Ca^{2+} , and be located correctly within



100 YEARS AGO

The September issue of the *American Naturalist* commences with an interesting paper, by Miss (or ? Mrs.) Sampson, on unusual modes of development among frogs and toads... Two species, for example, the one from West Africa and the other from Brazil, deposit their spawn in nests formed of leaves stuck together, the tadpoles moving in a mass of froth, recalling that of the cuckoo-spit insect. In both these instances the spawn is deposited in the neighbourhood of water, into which the tadpoles ultimately fall; but in a tree-frog from Rio, in which the eggs are likewise hatched in a frothy mass among leaves, the larvae actually die if they are put into water... Full reference is made to the mode of development in the Surinam toad, and also to that of the marsupial frogs, in which the young are hatched in a dorsal pouch. But perhaps the most extraordinary "nursery" arrangements in the entire group are those of the Chilean *Rhinoderma*, in which the tadpoles undergo their development in an enormous pouch on the throat of the male. From *Nature* 18 October 1900.

50 YEARS AGO

A paper by H. F. Finch "On a Periodic Fluctuation in the Length of the Day" gives a full account of the performance of a number of quartz-crystal clocks employed in the Greenwich Time Service... By applying the adopted current corrections to the standard clocks and using daily intercomparisons, it was possible to deduce the amount of correction for the clock controlling the time signal, prior to transmission. Although the method succeeded for a time, in the autumn of 1945 the observed clock corrections began to diverge from those predicted, the clocks appearing to have undergone an acceleration of rate, and a retardation in the period of rotation of the earth was suspected. In subsequent years, as more clocks became available, the same phenomenon was in evidence, but it was not until the spring of 1949 that a definite retardation among a number of clocks was noticed. The suspicion of a seasonal fluctuation in the earth's period of rotation led to an examination of the final corrections to the times of emission of the Rugby time signals which are under the control of the Greenwich Time Service... It appears that there is a variation of the order 0.01 sec and an accumulative effect in time of approximately 0.060 sec. From *Nature* 21 October 1950.

cells. BK channels in neurons occur close to plasma-membrane Ca^{2+} channels, and respond to the local Ca^{2+} signals they produce⁶. BK channels in smooth muscle cells are tuned (because of the $\beta 1$ subunit) to respond to the Ca^{2+} spark generated by ryanodine-sensitive channels on the sarcoplasmic reticulum. It may be that the BK and ryanodine-sensitive channels occur close together in smooth muscle. Brenner *et al.* speculate that the $\beta 1$ subunit could play a role in channel localization. In addition, BK channels — like several other channels⁷ — might be organized together with other proteins to form signalling complexes. As well as responding to Ca^{2+} (and voltage), BK channels are sensitive to modulation by phosphorylating enzymes, and a signalling complex might integrate the effects of these factors on the channel.

Brenner *et al.*'s work also adds to our

understanding of the mechanisms that regulate blood pressure, with possible benefit to patients suffering from hypertension. BK channels might merit investigation as potential targets for new drugs with which to treat this condition. And it will certainly be worth finding out whether the genes encoding components of BK channels contribute to its genetic basis in humans.

Nick Standen is in the Department of Cell Physiology and Pharmacology, University of Leicester, PO Box 138, Leicester LE1 9HN, UK.
e-mail: nbs@le.ac.uk

1. Brenner, R. *et al.* *Nature* **407**, 870–876 (2000).
2. Nelson, M. T., Patlak, J. B., Worley, J. F. & Standen, N. B. *Am. J. Physiol.* **259**, C3–C18 (1990).
3. Jaggar, J. H., Porter, V. A., Lederer, W. J. & Nelson, M. T. *Am. J. Physiol.* **278**, C235–C256 (2000).
4. Nelson, M. T. *et al.* *Science* **270**, 633–637 (1995).
5. Jiang, Z., Wallner, M., Meera, P. & Toro, L. *Genomics* **55**, 57–67 (1999).
6. Marrión, N. V. & Tavalin, S. J. *Nature* **395**, 900–905 (1998).
7. Pawson, T. & Scott, J. D. *Science* **278**, 2075–2080 (1997).

Geochemistry

Tracing the Earth's evolution

Mark Rehkämper

The noble metals provide unique clues to the early origins of our planet. But how they are distributed within the Earth — and what it all means — is the subject of intense debate. According to the most widely accepted model of the Earth's formation, the mantle received its supply of noble metals only after the Earth had already acquired about 99% of its present mass, during a final phase of bombardment by meteorites. But many rocks in the mantle have noble-metal abundances that are difficult to reconcile with such an extraterrestrial origin. On page 891 of this issue, Alard *et al.*¹ take a fresh look at this problem. Their results confirm that many rocks in the mantle have different noble-metal abundance signatures to meteorites. But there is a twist. They also provide conclusive evidence that these abundances cannot be used to argue against an extraterrestrial source for the noble metals in the mantle.

Piecing together the evolution of the Earth is difficult, because recent geological processes have erased most traces of our planet's early history. It is clear, however, that the primitive material that accreted to form the Earth was quickly differentiated into a silicate magma and dense metal phases. At an early stage, the metal sank to form the largely iron core that constitutes one third of the Earth's mass.

The noble metals osmium, iridium, platinum, gold, ruthenium, rhodium, palladium and rhenium are critical for deciphering how this happened because they display a strong affinity for metal magmas. In the laboratory,

these elements have a more than 10,000 times greater preference for metal liquids than for the silicate magmas from which the mantle formed^{2,3}. So more than 99% of the Earth's noble metals reside in the core, and only trivial amounts are expected to be in the mantle.

To characterize the noble-metal content of the Earth's mantle, geochemists analyse

peridotitic rocks, such as lherzolites, because these have compositions similar to that estimated for the Earth's primitive mantle. The first measurements of noble-metal concentrations in lherzolites, made about 20 years ago, revealed that the noble metals occur at concentrations about 100 times higher than would be expected if the mantle was in chemical equilibrium with the core. This is unexpected, because core formation strips the mantle of its noble metals.

Furthermore, the relative abundances of noble metals in the mantle (for example palladium/iridium, Pd/Ir) were roughly equal to those measured for chondritic meteorites, which are thought to represent the primitive material from which the Earth accreted^{4,5}. Among the many explanations offered for these observations, the 'late veneer' model has been the most successful. This model suggests that 0.5–1% of primitive chondritic material arrived at the Earth after core formation was completed⁶. The high concentrations of noble metals in chondrites mean that even the addition of a small amount of meteoritic material can account for the current abundances of noble metals in the mantle.

The 'late veneer' model has recently been challenged, as new (and more accurate) measurements of mantle peridotites identified noble-metal ratios that are difficult to reconcile with a primitive extraterrestrial source. Such measurements are extremely difficult because the concentration of noble elements is only a few parts per billion. So early geochemical studies were able to establish only that the noble metals in the mantle have roughly the same relative abun-

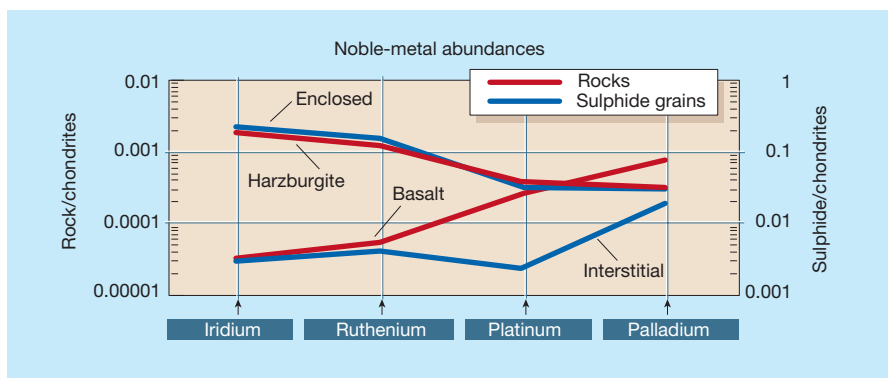


Figure 1 Noble-metal patterns that reflect melting processes rather than the Earth's early history. Comparison of the abundances of selected noble metals measured by Alard *et al.*¹ in sulphide grains from a primitive mantle rock (Australian lherzolite) and the abundances in whole rocks from the mantle (a residual East African harzburgite and a basalt from Kolbeinsey Ridge, Atlantic Ocean). The abundances are normalized to the concentrations found in chondritic meteorites. The high concentrations of the noble metals in sulphides show that sulphides dominate the budget of these elements in peridotites. One exception to this rule is platinum, which may enter other minor phases¹. The noble-metal signature of the sulphide enclosed in a silicate grain is almost identical to that of the harzburgite. With the exception of platinum, the noble metals in the interstitial sulphide found between the grains display a signature similar to that of basalt. So the two sulphides have the complementary patterns that are expected for a melt and its mantle residue. From these data, Alard *et al.*¹ argue that the unusual patterns of noble-metal abundances seen in mantle peridotites can be explained by ordinary melting processes.

dances as in chondrites. Nowadays, noble-metal abundances can be determined in mantle rocks with a precision and accuracy of about 10%.

Unexpected results were found for some peridotites, which had higher Pd/Ir ratios than the chondrites. These data demonstrate that some parts of the Earth's mantle are characterized by noble-metal signatures that are too complex to be simply of chondritic origin. From this, it was suggested that the noble metals in the mantle were derived from the accretion of less primitive meteorites or from an influx of core material rich in noble metals^{7,8}. Other studies proposed that the non-chondritic Pd/Ir ratios of peridotites could be due to ordinary rock-forming processes in the upper mantle⁹.

Alard *et al.*¹ now provide compelling evidence that mantle peridotites with non-chondritic Pd/Ir ratios can indeed be attributed to common processes in the upper mantle. The authors unravel the geochemistry of mantle rocks by studying individual mineral grains, rather than the gram-sized whole-rock samples used in previous studies. They use a laser-based microanalytical system to identify and analyse the individual micrometre-sized minerals that dominate the noble metals in mantle peridotites. Their results indicate that most of the noble metals in these rocks are concentrated in sulphide grains, which occur as two texturally different types. Some peridotites have only one kind of sulphide, whereas others display both varieties. The sulphides contain noble metals at a few parts per million, sufficient for precise measurement of concentration.

Alard and colleagues¹ discover that the two types of sulphide are also characterized by different, or rather complementary, signatures of noble-metal abundances. Sulphides enclosed in silicate mineral grains have low proportions of platinum and palladium relative to osmium, iridium and ruthenium. As such, they have noble-metal ratios similar to those observed in mantle residues such as harzburgites (Fig. 1). The authors interpret these sulphides as being residues from partial melting. The interstitial sulphides, which occur only along grain boundaries, have high Pd contents and low concentrations of osmium, iridium and ruthenium, similar to volcanic rocks such as basalts (Fig. 1). Alard *et al.* suggest that these sulphides crystallized from sulphide-containing fluids. Together, the sulphides have the complementary signatures expected for a melt and its residue.

The sulphide results demonstrate that melting processes can affect the noble-metal abundances of peridotites to produce both depleted and enriched noble-metal ratios. The finding that interstitial sulphides have higher Pd/Ir ratios than chondrites suggests that such enriched signatures are not inherited from core material but are related to

upper-mantle processes involving sulphide enrichment from melts. Moreover, the results indicate that the varying noble-metal ratios found in peridotites from different localities are due to ordinary rock-forming processes rather than spatial variations in the composition of the late meteoritic veneer.

The conclusion that mixing of depleted and enriched sulphides can generate (almost) chondritic noble-metal abundance patterns is disturbing, as it suggests that even peridotites with a complex rock-forming history can have primitive noble-metal signatures. The unambiguous interpretation of whole-rock noble-metal data therefore requires detailed information about the sulphide phases of the rock. But there are also lherzolites with chondrite-like noble-metal abundance patterns that do not show evidence of a complex rock-forming history.

Palaeontology

Mitrates on the move

Henry Gee

Mitrates are fossils that look like nothing alive today, and so defy anatomists to make sense of them. They also feature in R. P. S. Jefferies's controversial ideas about vertebrate evolution, to which he and colleagues now add in a paper in *Lethaia*¹. The evidence they describe is behavioural — inferred fossil tracks left by two specimens just before they died.

Mitrates are known from marine sediments laid down some 500–360 million years ago; they were usually just a few millimetres in length, with blobby 'heads' and long, segmented 'tails'. The key to their interpretation seems at first to lie in their armour plating, for mitrates had a skeleton made of plates, each grown from a single crystal of calcite. This kind of skeleton is characteristic of echinoderms, such as starfishes. However, the calcite skeleton is arguably the only feature that links mitrates with echinoderms². Mitrates lacked the familiar five-fold symmetry of extant echinoderms, and the presence of the characteristic echinoderm water-vascular system in mitrates is moot.

That this type of skeleton is nowadays restricted to echinoderms does not rule out the possibility that it was once found in other animals. Phylogenetic analysis should not be swayed by what are taken to be 'key' characters³ and force the past to conform to the particularities of the present. This reasoning is the basis of Jefferies's long-held contention that mitrates and some other fossil groups, also conventionally regarded as echinoderms, are in fact more closely related to chordates⁴ — the group that includes

It is likely that their noble-metal geochemistry has remained largely undisturbed and so still carries the primitive signature of the Earth's late meteorite bombardment. ■

Mark Rehkämper is at the Institute of Isotope Geology and Mineral Resources, ETH Zürich, CH-8092 Zürich, Switzerland.

e-mail: markr@erdw.ethz.ch

1. Alard, O., Griffin, W. L., Lorand, J. P., Jackson, S. E. & O'Reilly, S. Y. *Nature* **407**, 891–894 (2000).
2. Jones, J. H. & Drake, M. J. *Nature* **322**, 221–228 (1986).
3. Holzheid, A., Sylvester, P., O'Neill, H. St C., Rubie, D. C. & Palme, H. *Nature* **406**, 396–399 (2000).
4. Morgan, J. W., Wandless, G. A., Petrie, R. K. & Irving, A. J. *Tectonophysics* **75**, 47–67 (1981).
5. Mitchell, R. H. & Keays, R. R. *Geochim. Cosmochim. Acta* **45**, 2425–2442 (1981).
6. Chou, C.-L. *Proc. Lunar Planet. Sci. Conf. IX*, 219–230 (1978).
7. Pattou, L., Lorand, J. P. & Gros, M. *Nature* **379**, 712–715 (1996).
8. Snow, J. E. & Schmidt, G. *Nature* **391**, 166–169 (1998).
9. Rehkämper, M. *et al. Earth Planet. Sci. Lett.* **172**, 65–81 (1999).

vertebrates such as ourselves. For example, Jefferies suggests that perforations in the 'heads' of mitrates and the related cornutes are homologous to the structures known as pharyngeal slits found in hemichordates (acorn worms and allies) and chordates.

This view has not been upheld by a

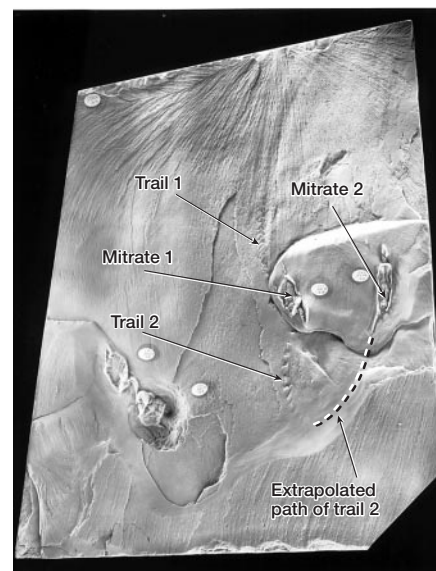


Figure 1 Fossil freeze-frame. These are the two strings of fossil marks that R. P. S. Jefferies and colleagues¹ have identified in 389-million-year-old slate. They interpret them as mitrate trails, and as constituting behavioural support for the idea that mitrates had chordate, rather than echinoderm, affinities. Two further mitrate fossils without trails appear at lower left. (Photograph courtesy of R. P. S. Jefferies.)

phylogenetic test⁵. Nevertheless, a molecular analysis has grouped echinoderms more closely with hemichordates than with chordates⁶; this is contrary to conventional understanding in which chordates and hemichordates form a group at the expense of echinoderms. This suggests that the common ancestor of all three groups had pharyngeal slits, a feature subsequently lost in the echinoderm lineage. It therefore constitutes indirect support for at least one of Jefferies' reconstructions of mitrate anatomy.

Further indirect evidence now comes from the description¹ of trails associated with two mitrates (*Rhenocystis latipedunculata*) preserved on a slab of slate 389 million years old (Fig. 1). The mitrates, living on the sea bed or just under it, seem to have been buried by a mudflow and were preserved along with the trails they had just made in the sea bed. Interpretation of the trails depends on how mitrates moved, which in turn depends on their anatomy.

The roughly lens-shaped mitrate 'head' has two distinct sides — one flat, made of large, pavement-like plates; the other convex, a patchwork of smaller plates. The end of the head opposite the 'tail' has a large orifice (mouth, anus or both, depending on whose work you read), ornamented at each corner by a spine that might have been moveable in life. The 'tail' is segmented and resembles, if superficially, the stalk or arm, respectively, of extant echinoderms such as crinoids (sea lilies) or ophiuroids (brittle stars).

There are three interpretations of mitrate orientation. In two of them, the animal is reconstructed with the 'head' convex side upwards. The third, that of Jefferies, has the 'head' with the flat side upwards. According to this view, the animal moved through soft sediment like a boat, displacing mud upwards and sideways as it travelled 'tail' first, using the 'tail' to hook itself through the sediment. This interpretation is supported by the disposition of ridges on the mitrate skeleton that resembles that of ridges on the shells of molluscs that live in seabed mud⁴.

The new work provides a fresh angle on Jefferies' argument that mitrates moved flat-side-up, 'tail'-first. One of the mitrates on the slab seems to have been buried while moving 'tail'-first, its trail being caused by lateral sweeping motions of one of the spines associated with the major orifice in the 'head'. The other trail is less well preserved but seems to have been caused by the motions of the 'tail' itself.

For various highly involved anatomical reasons, Jefferies' view of mitrates as chordates requires that mitrates were oriented with the flat side uppermost. This latest work does not, of itself, confirm Jefferies' assessment of mitrate anatomy. But along with other evidence⁶, it should mean that his ideas are revisited. The only proper test

will be phylogenetic, and it is true that the only credible, independent analysis so far suggests that mitrates have no particular connection with chordate evolution⁵. However, the real lesson is that extinct creatures can present us with a baffling mélange of characters⁷ that we would conventionally associate with different, extant groups. So it is possible that mitrates should properly be regarded as primitive, 'stem group' echinoderms³ — but echinoderms with features

that we would usually associate with chordates or hemichordates.

Henry Gee is a senior editor at Nature.

1. Sutcliffe, O. E., Südkamp, W. H. & Jefferies, R. P. S. *Lethaia* **33**, 1–12 (2000).
2. Smith, A. B. *Palaeontology* **27**, 431–459 (1984).
3. Budd, G. E. & Jensen, S. *Biol. Rev.* **75**, 253–295 (2000).
4. Jefferies, R. P. S. *The Ancestry of the Vertebrates* (British Museum, Natural History, London, 1986).
5. Peterson, K. J. *Lethaia* **28**, 25–38 (1995).
6. Cameron, C. B., Garey, J. R. & Swalla, B. J. *Proc. Natl Acad. Sci. USA* **97**, 4469–4474 (2000).
7. Clack, J. A. *Nature* **394**, 66–69 (1998).

Global change

A world in transition...

Harry Elderfield

Decrease in greenhouse-gas levels. Rapid increase in the size of continental ice sheets. Abrupt climate shift. A switch in ocean circulation and mass extinctions of marine creatures. Here we are at the Eocene–Oligocene boundary, about 33.7 million years ago, a crucial time in the Earth's history and of especial interest because it marked a dramatic move from a 'greenhouse world' to an 'icehouse world'. Temperature change is often cited as the prime suspect in causing the extinctions, but the idea is controversial.

A new twist on events comes from the work by Ivany and colleagues¹ that appears on page 887 of this issue. They show that although inferred temperatures did not change across the boundary, seasonality did: summer temperatures remained the same

but winters became cooler. Was this the cause of the mass marine extinctions? Did cold winters kill millions?

Central to this debate is the estimation of temperature using the oxygen-isotope composition of biologically generated carbonate in the oceans. The $^{18}\text{O}/^{16}\text{O}$ ratio of the calcium carbonate directly reflects the $^{18}\text{O}/^{16}\text{O}$ ratio in the sea water from which it precipitates but with an offset that is a well-defined function of temperature. Most previous studies have used the remains of foraminifera, marine unicellular organisms that lay down chambers of calcium carbonate. The results^{2–4} give an upper cooling limit of 3–4 °C in the deep oceans across a period of only about 50,000 years at the Eocene–Oligocene boundary. If we assume that, as today, formation of the oceans' deep waters

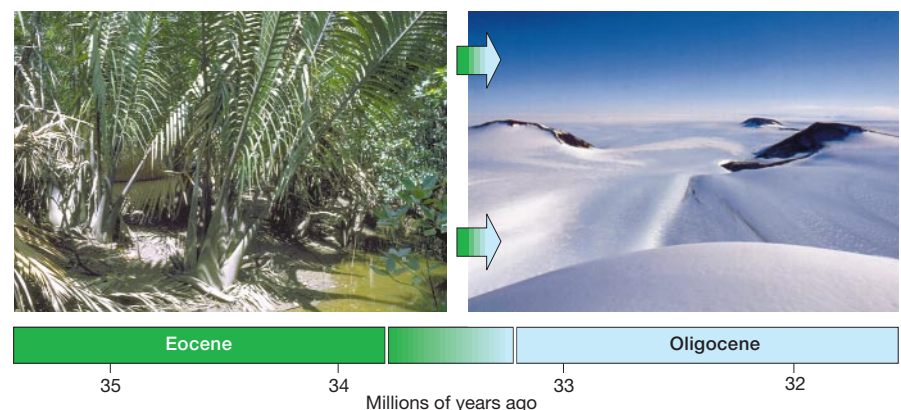


Figure 1 From greenhouse to icehouse, across the Eocene–Oligocene boundary. Remains of tropical plants such as those on the left are found in many lower Eocene deposits from high latitudes; in the Oligocene, Antarctic glaciation developed and the seas fell to their lowest levels of the past 100 million years. A possible course of events is that tectonic movements led to widening of the oceanic passages, isolating Antarctica and allowing development of the circumpolar current. This provided a moisture source for snow precipitation and led to the initiation of a large-scale Antarctic ice sheet. At the same time, a gradual long-term lowering of atmospheric CO_2 , perhaps driven by changes in the intensity of weathering (which uses up CO_2 in the formation of carbonic acid, the weathering agent of rocks), diminished the greenhouse effect. As the temperature gradient steepened between the poles and the tropics, oceanic and atmospheric circulation intensified. The system flipped from greenhouse to icehouse. The changes in seasonality identified by Ivany *et al.*¹ may have resulted from the oceanic perturbation and in turn have led to the marine extinctions.

occurs by sinking of sea water at high latitudes, then this also represents cooling of polar surface waters. In contrast — and here is the ambiguity — no clear indication has been found of sea-surface temperature change at low latitudes at the boundary.

Ivany *et al.* have used fish otoliths ('ear stones') collected from an Eocene–Oligocene boundary section in the Gulf Coastal Plain of the southern United States; the section represents original conditions of about 50 m water depth on the continental shelf. The work's novelty lies in microsampling of the otolith calcium carbonate, which, via the $^{18}\text{O}/^{16}\text{O}$ ratios, provides a record both of the average temperature during the fish's lifetime and of the seasonal range in temperature. About 30 microsamples were drilled from the 3–4-mm-thick otoliths. The otolith powders were then analysed for oxygen isotopes, each sample representing less than two weeks of growth from between 28 million and 36 million years ago.

This impressive technical feat has produced two notable results. First, wide seasonal ranges in temperatures of up to more than 10 °C are recorded by each otolith during its growth, and there is no clear evidence that the average temperature changed across the boundary. Perhaps this is not too surprising because of the low latitude of the study area. Second, and more significantly, whereas summer temperatures did not change, winter temperatures fell by as much as 4 °C. The authors wonder whether it is this change in seasonality, rather than a drop in mean annual temperature, that was the cause of the extinctions.

Ivany *et al.* themselves discuss some caveats about their data and interpretations. The conversion of oxygen-isotope ratios in calcium carbonate to a temperature estimate requires knowledge of the oxygen-isotope composition of sea water in which the calcification occurred. This is tricky. During evaporation at the sea surface, the lighter ^{16}O isotope in water evaporates faster than the heavier ^{18}O . As a result, isotopically depleted water vapour is transported away from the sub tropics towards the poles, generating oceanic gradients in $^{18}\text{O}/^{16}\text{O}$ ratios and producing snow and ice with very low levels of ^{18}O . The overall result is that the $^{18}\text{O}/^{16}\text{O}$ ratio of the oceans as a whole rises with increased continental ice volume.

The great advantage of Ivany and colleagues' work is that the seasonality they calculate is independent of any change in seawater $^{18}\text{O}/^{16}\text{O}$. Only the absolute values of temperature are affected: in the calculation of temperature difference, the seawater $^{18}\text{O}/^{16}\text{O}$ term cancels out so their conclusion is robust irrespective of how seawater $^{18}\text{O}/^{16}\text{O}$ changed through time. But there is a complication. Ivany *et al.* used values for the $^{18}\text{O}/^{16}\text{O}$ of seawater from the deep-sea foraminifera work I mentioned earlier^{2–4},

which, recognizing that changes in ice volume must have occurred, attributed part of the change in foraminiferal $^{18}\text{O}/^{16}\text{O}$ across the boundary to temperature and the remainder to changes in global oceanic $^{18}\text{O}/^{16}\text{O}$.

An alternative approach⁵ to estimating temperatures involves using the magnesium content of the foraminifera (I'll skip the details). From these data it seems that the first major accumulation of Antarctic ice was not accompanied by a decrease in deep-sea temperatures. In other words, the surmise here is that all of the oxygen-isotope shift was caused by changes in ice volume. When Ivany *et al.* use seawater oxygen-isotope values based on the magnesium work, the temperature difference remains the same but it inverts their results: winters remained constant but summers became warmer. This seems unlikely. I think the problem lies in the difficulty of defining seawater $^{18}\text{O}/^{16}\text{O}$ at their study sites. The $^{18}\text{O}/^{16}\text{O}$ ratio of the oceans as a whole responds to continental ice volume but there are also wide ranges in seawater $^{18}\text{O}/^{16}\text{O}$ from place to place for the reasons I touched on earlier. We simply do not know enough about the hydrological cycle at that time.

How does this result fit into the picture for the end of the greenhouse world? The title of this article is a truncated form of a line taken from Prothero⁶: "A world in transition... then all hell breaks loose." It invokes a transitional period⁴ at the end of the warm, mainly ice-free Eocene, followed by rapid climate deterioration in the earliest

Oligocene to glacial conditions. There may have been two triggers (Fig. 1): first, widening of the oceanic passages that isolated Antarctica, which allowed the circumpolar current to form; second, a long-term decrease in atmospheric CO_2 . But the sequence of events associated with and following the transition remain poorly understood. There seems no doubt that a massive change in the pattern of deep-ocean circulation occurred. Ivany *et al.* argue that the increased seasonality they propose is more consistent with climate change driven jointly by the oceans and atmosphere than by a reduction in greenhouse gases alone.

Their study is a fascinating example of high-resolution analysis of climate change in that they identify annual temperature shifts in the oceans 30 million years ago. But it offers only relatively coarse resolution of precisely when seasonality changed. Addressing this issue will be one of the next steps to take using the otolith technique. ■

Harry Elderfield is in the Department of Earth Sciences, University of Cambridge, Downing Street, Cambridge CB2 3EQ, UK.

e-mail: he101@esc.cam.ac.uk

1. Ivany, L. C., Patterson, W. P. & Lohmann, K. C. *Nature* **407**, 887–890 (2000).
2. Miller, K. G. *et al.* *Paleoceanography* **2**, 1–18 (1987).
3. Zachos, J. C., Stott, L. D. & Lohmann, K. C. *Paleoceanography* **9**, 353–387 (1994).
4. Zachos, J. C., Quinn, T. M. & Salamy, K. A. *Paleoceanography* **11**, 251–266 (1996).
5. Lear, C. H., Elderfield, H. & Wilson, P. A. *Science* **287**, 269–272 (2000).
6. Prothero, D. R. *The Eocene–Oligocene Transition* (Columbia Univ. Press, New York, 1994).

Gene regulation

One man's food

Mitchell A. Lazar

Since early in our evolution, we have been creating unnatural molecules — from cooked foods to food additives, drugs to building materials — with the intention of improving our environment and well-being. These 'xenobiotics' can be ingested in food, air or water, either intentionally (in the case of food or drugs) or unintentionally. Hydrophobic xenobiotics are converted into fat-soluble molecules in the liver by enzymes of the 'cytochrome P450' family¹. This process detoxifies harmful xenobiotics, but has drawbacks — it can deactivate therapeutic drugs and convert xenobiotics that are not intrinsically dangerous into toxins or carcinogens. From a therapeutic standpoint, there is another problem: the activation of the conversion enzymes by one xenobiotic often increases the metabolism of other compounds, making it hard to predict what effects drugs may have.

Not surprisingly, there is a lot of interest in understanding how xenobiotics kick-start the conversion process. Wei *et al.* and Xie *et al.*, writing on page 920 of this issue² and in a recent issue of *Nature*³, respectively, have been probing this process, with some interesting outcomes. One of these is the production of a 'humanized' mouse model that might be useful for testing the effects of drugs.

Multicellular organisms are the product of the interplay between their genes and their environment. One way in which the environment influences gene expression is through the proteins of the nuclear hormone-receptor superfamily⁴. Hormones that bind to these receptors are small, fat-soluble molecules secreted by ductless glands such as the thyroid, adrenals and gonads. Members of this receptor superfamily also allow vitamins A and D — supplied in food — and the

Daedalus

David Jones

David Jones, author of the Daedalus column, is indisposed.

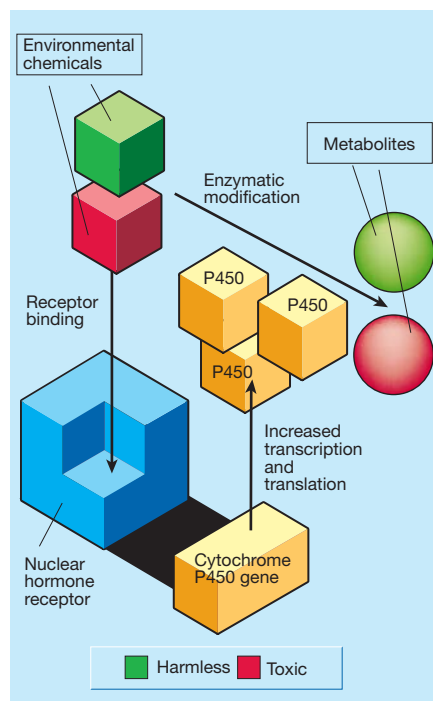


Figure 1 Unravelling the basis of drug toxicity. When ingested in food, water or air, some unnatural molecules ('xenobiotics'), such as drugs, bind to nuclear hormone receptors in liver cells. These receptors can then activate the expression of cytochrome P450 genes. The cytochrome P450 enzymes then convert the xenobiotics into other forms. This process can be a double-edged sword: it can lead to the detoxification of harmful compounds, but can also convert harmless xenobiotics into toxins or carcinogens. The activation of the conversion enzymes by one xenobiotic can also enhance the metabolism of other xenobiotics. Wei *et al.*² have identified the nuclear hormone receptor (called CAR) that is responsible for the increased toxicity of cocaine in the presence of phenobarbital. Xie *et al.*³, meanwhile, have shown that the receptors themselves are responsible for the species specificity in the activation of different cytochrome P450 enzymes. Xie *et al.* have also generated mice that show a 'human' response to certain drugs.

products of their metabolism to regulate gene expression. These and other ligands for the nuclear hormone receptors are all naturally occurring 'endobiotics', which help to maintain the internal balance of the organism in the face of changes in the external and internal environments.

Many xenobiotics induce the expression of one or more of the cytochrome P450 genes, and hence bring about their own metabolism. Over the past few years it has become clear that this effect is mediated through nuclear hormone receptors (Fig. 1). 'Orphan' nuclear hormone receptors (those for which the natural ligand is not yet known) can bind xenobiotics and activate the expression of cytochrome P450 genes¹.

And studies of mice engineered to lack a nuclear hormone receptor — specifically PPAR- α ⁵ and, now, CAR² and PXR/SXR³ — confirm that these proteins are responsible for the induction of cytochrome P450 genes by xenobiotics *in vivo*.

Wei *et al.*'s 'gene-knockout' studies² also provide insight into the problem of drug toxicity and drug interactions. Cocaine kills liver cells, and this toxicity is exacerbated by the anticonvulsant drug phenobarbital⁶. Wei *et al.* discover why. They find that phenobarbital activates the nuclear hormone receptor CAR. Next, CAR enhances the expression of a gene encoding a class 2 cytochrome P450 enzyme. The cytochrome P450 enzymes detoxify phenobarbital, but the process has the unhappy side effect of converting cocaine to a more toxic form.

Wei *et al.*² and Xie *et al.*³ also offer some answers to another question: might there be a link between the xenobiotics that bind to a given nuclear receptor, and the substrate specificity of the class of cytochrome P450 enzymes that tend to be regulated by that receptor? Consider the metabolism of the anaesthetic zoxazolamine. Xie *et al.* found that mice engineered to have a higher than normal amount of active PXR/SXR were protected from paralysis induced by zoxazolamine. The mice also have increased levels of class 3 cytochrome P450 enzymes. These results suggest a direct connection between the activation of the nuclear hormone receptor PXR/SXR by zoxazolamine, an increase in class 3 cytochrome P450 enzymes, and the detoxification of this anaesthetic.

But the situation is more complex than this. Wei *et al.* found that phenobarbital protects normal but not CAR-deficient mice from zoxazolamine-induced paralysis, suggesting that CAR and its cytochrome P450 target(s) are also required for the inactivation of zoxazolamine. And mice lacking one of the class 1 cytochrome P450 genes (which is activated by yet another class of xenobiotic receptor⁷) show increased sensitivity to zoxazolamine. It will be interesting to know whether mice that lack more than one of the nuclear hormone receptors are even more sensitive to zoxazolamine and other toxins.

The nuclear hormone receptors for xenobiotics have low affinity for a great many structurally unrelated compounds. This is consistent with their role in detecting a wide variety of foreign molecules. But it does not necessarily mean there is no specificity in the process. Some xenobiotics activate the expression of the conversion enzymes in a species-specific way⁸. For example, the antibiotic rifampicin activates class 3 cytochrome P450 enzymes in humans, but not in mice. Xie *et al.*³ suggest that the origin of this example of species specificity lies in the nuclear hormone receptor PXR/SXR. The authors expressed human PXR/SXR in

the livers of mice engineered to lack the mouse version of this receptor. They found that the mice showed a 'human' response to xenobiotics such as rifampicin. This 'humanized' mouse model could allow more accurate predictions of how toxic certain drugs are, and how they interact with other drugs, before clinical trials are needed. The range of xenobiotics that activate other nuclear hormone receptors is also species specific^{9,10}, so there may be potential in developing mouse models containing the human versions of these proteins, too.

The existence of the nuclear hormone receptors suggests that they provided a survival advantage during evolution. The differences between these receptors in humans and mice probably reflect their need for protection from distinct environmental challenges. But even within a single species, the conversion of xenobiotics by nuclear hormone receptors can be a double-edged sword. In humans, the environment is an important cause of cancers¹¹, and some xenobiotics are major contributors to this process.

In this industrial age, the role of nuclear hormone receptors in human health and disease is very complex. So what can we do with this new knowledge? We might save lives by manipulating specific human receptors so they detoxify particular environmental carcinogens. But such attempts could be fatal if they create new and more dangerous toxins. Which xenobiotics do we want to metabolize, and how can we do it selectively? The Roman poet Lucretius anticipated this modern dilemma in his warning: "What is food to one man may be fierce poison to others"¹².

Mitchell A. Lazar is in the Division of Endocrinology, Diabetes and Metabolism, Departments of Medicine and Genetics and The Penn Diabetes Center, University of Pennsylvania School of Medicine, Philadelphia, Pennsylvania 19104, USA.

e-mail: lazar@mail.med.upenn.edu

1. Waxman, D. J. *Arch. Biochem. Biophys.* **369**, 11–23 (1999).
2. Wei, P., Zhang, J., Egan-Hafley, M., Liang, S. & Moore, D. D. *Nature* **407**, 920–923 (2000).
3. Xie, W. *et al.* *Nature* **406**, 435–439 (2000).
4. Mangelsdorf, D. J. *et al.* *Cell* **83**, 835–839 (1995).
5. Lee, S. S. T. *et al.* *Mol. Cell. Biol.* **15**, 3012–3022 (1995).
6. Boelsterli, U. A. & Goldin, C. *Arch. Toxicol.* **65**, 351–360 (1991).
7. Okino, S. T. & Whitlock, J. P. *Vitam. Horm.* **59**, 241–264 (2000).
8. Jones, S. A. *et al.* *Mol. Endocrinol.* **14**, 27–39 (2000).
9. Moore, L. B. *et al.* *J. Biol. Chem.* **275**, 15122–15127 (2000).
10. Holden, P. R. & Tugwood, J. D. *J. Mol. Endocrinol.* **22**, 1–8 (1999).
11. Lichtenstein, P. *et al.* *N. Eng. J. Med.* **343**, 78–85 (2000).
12. Bartlett, J. (ed.) *Bartlett's Familiar Quotations* 9th edn (Little, Brown, Boston, 1901).

Obituary

Robert H. Abeles (1926–2000)

Scientific discoveries often arrive when a hypothesis fails. Such was the experience of the enzymologist Robert Abeles, who died on 18 June 2000.

Abeles was convinced that enzymatic reactions could be described by the principles of organic chemistry. In particular, he believed that the transfer of a hydrogen from one carbon to its neighbour (a 1,2-hydride shift) provides a simple mechanism for biological oxidation reactions — the metabolic equivalent of combustion, which constitute a cell's primary source of energy. Abeles set out to verify the hypothesis with an investigation of diol dehydrase, a bacterial enzyme. Alas, the reaction did not involve a classic 1,2 hydride shift. Instead, he revealed a reaction previously beyond the imagination: cleavage of a carbon–cobalt bond to produce a radical. This reaction is the key to the chemistry of vitamin B₁₂, a molecule whose importance had long been recognized if not understood. It was the first time that biochemistry had anticipated chemistry.

This was 1958, when Abeles had just left a postdoctoral position in Frank Westheimer's laboratory at Harvard to work at Ohio State University. He observed that diol dehydrase required a light-sensitive cofactor similar to the vitamin B₁₂ coenzyme that had just been discovered by H. A. Barker at Berkeley. Dorothy Hodgkin simultaneously found by X-ray crystallography that the coenzyme contained an unprecedented carbon–cobalt covalent bond. However, many biochemists believed that the coenzyme was simply a bystander to the actual chemical transformation.

Abeles quickly demonstrated that the carbon–cobalt bond was reversibly cleaved during the diol dehydrase reaction. In 1964, he took up a post at Brandeis University, where he remained for the rest of his career. There he further showed that coenzyme vitamin B₁₂ acts as a radical initiator. Radicals are quite promiscuous in their reactivity and renowned for undergoing chain reactions. In contrast, enzymatic reactions involving vitamin B₁₂ are tightly controlled radical transformations. The surprise of this finding is difficult to comprehend today, when the ability of enzymes to tame reactive intermediates is largely taken for granted.

Abeles' work relied on the confluence of chemistry and biology long before the



A founder of modern enzymology

interface of these two disciplines came into vogue. He invariably chose to investigate reactions that posed mechanistic puzzles to organic chemists, and his contributions include many 'firsts'. He was the first to show that carbanions are early intermediates in flavoenzyme reactions; the first to find non-amino-acid residues in the active sites of enzymes; and among the first to demonstrate a covalent enzyme–substrate intermediate.

His imaginative and elegant experiments defined the mechanisms of enzymes in every reaction class, perhaps more than any other scientist. Abeles sought apparent biochemical contradictions, knowing that the resolution of such paradoxes would reveal new principles. For example, he recognized that a cell's demand for polyamines, compounds with a crucial although undefined role in cell physiology, must generate huge quantities of the by-product methylthioadenosine, a potent inhibitor of polyamine synthesis. This insight led him to unravel the pathway that converts methylthioadenosine to methionine; and then to the discovery of an enzymatic reaction that produces carbon monoxide, an internal regulator of blood-vessel relaxation.

In the early 1970s, Abeles started to concentrate on the rational design of small-molecule inhibitors of enzymes, anticipating the 'chemical genetics' movement of today. Such inhibitors are the foundation of drug discovery. Coupling enormous creativity with deep chemical knowledge, Abeles produced

potent inhibitors to proteases, carboxylases, oxidases, isomerases, dehydrogenases — again, to virtually every class of enzymatic reaction.

One especially notable achievement was his design of the first transition-state analogue (the transition state is the most unstable point in a chemical transformation, where bonds are breaking and forming). In this work, Abeles recognized the planar geometry of pyrrole-2-carboxylate in the planar transition state of proline racemase catalysis. Linus Pauling had proposed that the catalytic power of enzymes derives from the ability to bind transition states with high affinity. A compound that mimics the transition state will also bind with high affinity, and pyrrole-2-carboxylate indeed proved to be a potent inhibitor of proline racemase.

Abeles may be best known for developing 'suicide', or mechanism-based, inhibitors. These compounds undergo the initial steps of catalysis, which unmask dormant reactive groups, and the enzyme catalyses its own demise. As well as providing clever probes of enzyme mechanism, such inactivators promise the exceptional specificity necessary for drug development.

The first encounter with Bob Abeles was usually a transforming experience. Abeles had a magnetic speaking quality and a unique ability to stimulate others to think creatively. Colleagues envied his ability to strip a problem to its essence, and students marvelled at his uncanny scientific instincts. Together with William P. Jencks, he trained three generations of Brandeis graduate students in the rigours of mechanistic enzymology, establishing Brandeis as a Mecca of the field.

Perhaps Bob Abeles' greatest gift was his absorption in science, which may have enabled him to cope with the medical problems that came to plague him. In the 1970s he endured drug and radiation therapy for Hodgkin's disease. In the diagnostic biopsy, he suffered a severed vocal chord, which sadly affected his wonderfully distinctive voice. For the last ten years he also had Parkinson's disease. But through it all he never lost his scientific acuity or venturesome spirit.

Lizbeth Hedstrom and Perry A. Frey

Lizbeth Hedstrom is in the Department of Biochemistry, MS 009, Brandeis University, 415 South Street, Waltham, Massachusetts 02454-9110, USA.

e-mail: hedstrom@brandeis.edu

Perry A. Frey is in Department of Biochemistry, University of Wisconsin-Madison, 1710 University Avenue, Madison, Wisconsin 53705-4098, USA.

e-mail: frey@enzyme.wisc.edu

Genital damage, kicking and early death

The battle of the sexes takes a sinister turn in the bean weevil.

Because the costs and benefits of polygamy differ for males and females, copulation is not always a cooperative venture between the sexes¹. Sperm competition² can build on this asymmetry, producing male traits that harm females^{3,4}, thereby generating coevolutionary arms races between the sexes⁵. We have found that the male genitalia of the bean weevil *Callosobruchus maculatus* damage the female genitalia, and that females act to reduce the extent of this damage. We propose that these functionally diametric sexual traits form the basis of reproductive conflict.

The intromittent organ of male *C. maculatus* bears strongly sclerotized spines (Fig. 1) which are everted in the central region of the female's genital tract during copulation. Male *C. maculatus* do not remove the sperm of other males as a mechanism of sperm competition⁶, so it is unlikely that the spines

are used for this, as they are in other insect species⁷.

Examining the *in situ* genitalia of copulating pairs, we found that these sclerotized spines penetrated the cuticular lining of the female's genital tract (Fig. 2a). Once-mated females had repaired punctures (melanized haemocyte plugs) in the endocuticular lining of their genital tracts 16 hours after copulation (Fig. 2b) which were absent in virgins (Fig. 2c). (In virgins, the total area of melanized 'wound' tissue was $0 \pm 0 \mu\text{m}^2$, $n=20$; in once-mated females, the mean extent of repair was $112.75 \pm 69.6 \mu\text{m}^2$, $n=20$; $U=400$, $P<0.0001$.) The spines on the intromittent organ of male *C. maculatus* therefore damage the female genital tract

during copulation, a phenomenon that may occur in other insects⁸.

Female *C. maculatus* repeatedly kick males towards the end of copulation⁶. We prevented females from kicking their mates by ablating the kicking legs, and then compared their mean copulation duration with that of a treatment control group (ablated non-kicking legs) and an unmanipulated control group. Non-kicking females copulated for significantly longer (866.9 ± 1.07 s) than treatment control (501.3 ± 1.06 s) and unmanipulated control (588.8 ± 1.08 s) groups (log-transformed data, analysis of variance, $F=14.16$, d.f. = 2,69, $P<0.001$; Tukey, $P<0.001$ for each comparison between non-kicking and control groups;



Figure 1 Scanning electron micrograph of the male bean weevil's intromittent organ showing the arrangement of sclerotized spines on the tip. The spines unfold in the region of the female genital tract shown in Fig. 2. The intromittent organ is about 800 μm long. Micrograph by Andrew Syred, Microscopix.

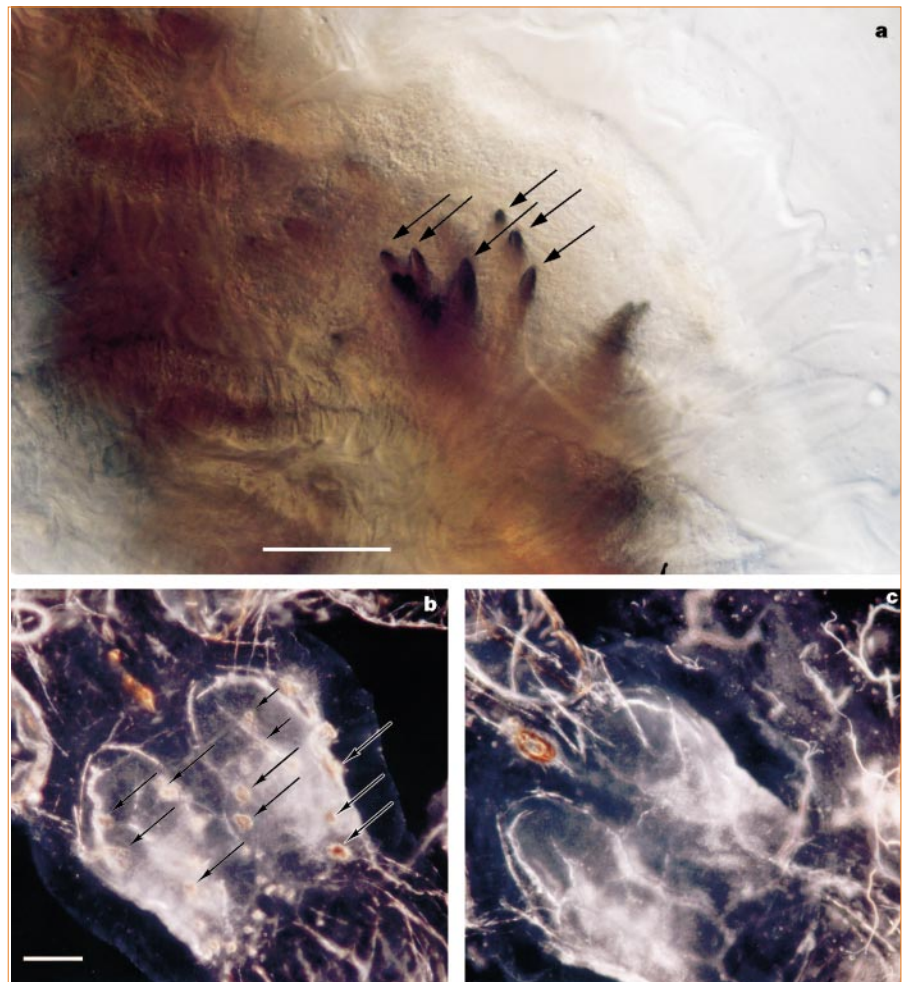


Figure 2 The anatomy of reproduction in *Callosobruchus maculatus*. **a**, Normarsky micrograph of the epithelial surface of the female genital tract containing the male intromittent organ: the specimen was dissected from a copulating pair snap-frozen in liquid nitrogen. After freezing, the *in copula* genitalia were removed and examined with a compound microscope (Leitz Diaplan). The photomicrographic image is focused on the epithelial surface of the female's genital tract: the arrows indicate the tips of spines that are in, and have passed through, this focal plane and so have penetrated the cuticular lining of the female's genital tract. **b**, A dark-field micrograph of the distal portion of a female genital tract 16 h after a single copulation. The arrows indicate melanized haemocyte plugs in the endocuticle. Anterior is towards the top left-hand corner in all micrographs. Scale bars, 30 μm . **c**, A dark-field micrograph of a dissection of the distal portion of a virgin female's genital tract.

control group means did not differ).

Mate-kicking by females therefore shortens copulation. Moreover, females that kicked during copulation sustained significantly less damage to their genital tracts ($147.06 \pm 84.9 \mu\text{m}^2$) compared with females that were unable to kick ($272.4 \pm 142 \mu\text{m}^2$) (log-transformed data, $t = -2.52$, d.f. = 38, $P < 0.02$).

If genital-tract damage is costly for females, there is probably a reproductive conflict over the frequency of copulation in *C. maculatus*. To determine whether copulation frequency has a life-history cost, we measured the longevity of singly and doubly mated females that were maintained in isolation with a limited number of oviposition sites to control egg-production costs (there was no difference in fecundity between treatments; $t = 0.009$, d.f. = 89, $P > 0.05$). Doubly mated females died significantly younger than singly mated females (log-rank Mantel-Cox, $\chi^2 = 5.42$, d.f. = 1, $P < 0.02$), indicating that remating reduces longevity, although reduced longevity does not necessarily result in reduced female fitness. The higher death rate of remated females could be due to genital damage incurred during copulation.

There are at least two, non-exclusive ways in which genital wounding could increase the fitness of male *C. maculatus*. First, it may reduce the risk of sperm competition⁹ if females postpone remating to delay receiving more damage to their genitalia. This may increase the number of eggs fertilized by the last male to mate. Second, genital damage may increase immediate oviposition rates because females perceive genital damage as a threat to survival and invest more in current reproduction; also, males transfer compounds that increase the rate of female egg-laying (as occurs in the fruitfly *Drosophila melanogaster*¹⁰) and which might enter the female's haemolymph rapidly through the genital punctures. An extended interval between female matings and/or increased oviposition rates will benefit the last male to copulate, as he will achieve sperm precedence⁶.

An antagonistic male trait — the transfer of ejaculatory compounds that alter female behaviour and reduce female longevity — has been documented in *D. melanogaster*^{3,10}, but it is unclear how female *D. melanogaster* respond to the selection pressure arising from the resultant reproductive conflict. Female *C. maculatus* seem to have responded to male-induced genital wounding by kicking their partner during copulation. This behaviour reduces the damage inflicted on their genital tract, and may therefore represent a counteradaptation to a potentially antagonistic male reproductive trait.

Helen S. Crudgington, Mike T. Siva-Jothy

Department of Animal & Plant Sciences,

University of Sheffield, Sheffield S10 2TN, UK

e-mail: m.siva-jothy@sheffield.ac.uk

1. Bateman, A. J. *Heredity* **2**, 349–368 (1948).
2. Parker, G. A. *Biol. Rev.* **45**, 525–567 (1970).
3. Chapman, T., Liddle, L. F., Kalb, J. M., Wolfner, M. F. & Partridge, L. *Nature* **373**, 241–244 (1995).
4. Arnqvist, G. & Rowe, L. *Proc. R. Soc. Lond. B* **261**, 123–127 (1995).

5. Rice, W. R. *Nature* **381**, 232–234 (1996).
6. Eady, P. *Ecol. Entomol.* **19**, 11–16 (1994).
7. Waage, J. K. *Science* **203**, 916–918 (1979).
8. Merritt, D. J. *Zoomorphology* **108**, 359–366 (1989).
9. Simmons, L. W. & Siva-Jothy, M. T. in *Sperm Competition and Sexual Selection* (eds Birkhead, T. R. & Møller, A. P.) 341–432 (Academic, London, 1998).
10. Baumann, H. J. *Insect Physiol.* **20**, 2347–2363 (1974).

Environmental chemistry

Is acidification still an ecological threat?

There has been a significant reduction in anthropogenic acid deposition in Europe and North America, and now we need to gauge the rate and extent of ecosystem recovery. Stoddard *et al.*¹ have reported a widespread aquatic recovery from acidification in European ecosystems in response to a fall in sulphate deposition. But many sites in central Europe are showing a significant delay in aquatic recovery from acidification, or even no recovery at all, and only some of them show biological recovery of waters or a recovery from soil acidification. Ecosystem management still needs to consider the consequences of acidification.

Results obtained in Europe from monitoring network programmes such as ICP-waters¹ (for International Cooperative

Programme on Assessment and Monitoring of Acidification of Rivers and Lakes) may bias overall conclusions because these programmes tend to focus on areas with high sensitivity to acid input. This high sensitivity is caused primarily by a low storage capacity for sulphate and protons (that is, soils that are characterized by all or a combination of: shallow post-glacial soil development, sandy soil texture, a high humus content). Data from soils with a high storage capacity for sulphate and protons indicate that reversing water acidification can take decades, because the release of previously stored sulphate causes cation leaching and acidification of deeper soil layers and waters.

Modelling by linear-regression analysis shows that the pH and alkalinity² of the streams at the German catchments Lehenbach, in Fichtelgebirge, northeastern Bavaria, and Lange Bramke, in the Harz mountains, Lower Saxony, are not increasing — even though sulphate and

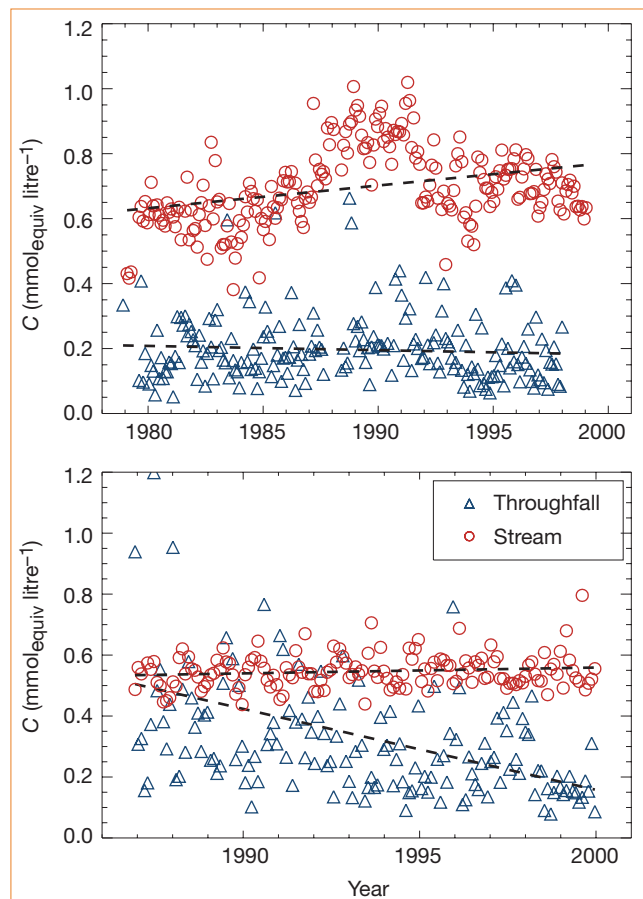


Figure 1 Nutrient cation concentration (where *C* represents the sum of the Ca^{2+} , Mg^{2+} , K^{+} and Na^{+} ion concentrations) in throughfall (precipitation below forest canopy; triangles) and streams (circles) at the catchments Lange Bramke (top) and Lehenbach (bottom) in Germany. Black dashed lines represent a linear regression and indicate the overall trend in *C* over the study period.

proton concentrations in throughfall and precipitation have fallen by 70% in the past ten years — indicating that there has been no aquatic recovery. Studies covering other regions of Germany (Solling³; Spessart, Frankenwald, Fichtelgebirge, Bayerischer Wald⁴; Schwarzwald⁵) or the Czech Republic⁶ have reported a continuing acidity and low pH in streams, or even an increase in sulphate concentrations in groundwater^{5,7}. This delayed recovery has been attributed to a release of previously stored soil sulphate in all these cases.

One of the main problems associated with soil and water acidification is the leaching of nutrient cations (Ca^{2+} , Mg^{2+} , K^+ , Na^+) from soils. Owing to a substantial reduction in dust emissions over the past 20 years, nutrient cation concentrations in precipitation have declined in some regions (Fig. 1, triangles). But in the streams of the German catchments studied here, these did not decline over the monitoring period (Fig. 1, circles). This was because the continued high sulphate and nitrate in streams cause high concentrations of cations. Although increased concentrations of nutrient cations in streams are usually taken as an indicator of water recovery^{1,8}, they result in a long-term export of nutrients and may point to a continuing (or even increasing) net acidification of soils. An alkalinity increase in streams, as reported by Stoddard *et al.*¹, may not necessarily indicate a reversal of soil acidification or biological recovery of the affected streams.

Stoddard *et al.*¹ suggest that a larger decrease in sulphur deposition and/or a longer response time may be required for a widespread recovery to occur in North America, which they claim lags behind the Nordic countries. But the decrease in sulphate and proton deposition started ten years earlier in North America than in Europe (beginning in the late 1960s and early 1970s) and deposition was less⁹. Although the Scandinavian countries began reducing their emissions early on, they continued to receive high deposition from central Europe until the early 1980s. Furthermore, a decrease in sulphate deposition would only increase alkalinity in streams as long as the present nutrient cation supply from mineral weathering is high enough to match both the sulphate and the nitrate fluxes in streams. However, no significant decrease in nitrogen deposition or nitrate concentrations in streams can be expected in the near future in either North America or Europe.

Any delay in recovery from acidification has important implications. Forest management still calls for frequent liming to avoid depletion of soils and a deficiency of nutrients for trees; water authorities need to consider a delay of decades in the recovery of streams (and thus drinking-water supplies)

draining from soils with a high sorption capacity for sulphate. Nitrogen deposition must also be reduced to limit both cation depletion from soils and the further acidification of waters.

C. Alewell*, **B. Manderscheid†**,
H. Meesenburg‡, **J. Bittersohl‡**

*BITÖK, University of Bayreuth,

95440 Bayreuth, Germany

e-mail: christine.alewell@bitok.uni-bayreuth.de

†Forest Research Station of Lower Saxony,

37079 Göttingen, Germany

‡Bavarian State Office for Water Management,

80636 Munich, Germany

1. Stoddard, J. L. *et al.* *Nature* **401**, 575–578 (1999).

2. Reuss, J. O. & Johnson, D. W. (eds) *Acidic Deposition and the Acidification of Soils and Waters* (Springer, New York, 1986).

3. Alewell, C., Bredemeier, M., Matzner, E. & Blanck, K. *J. Environ. Qual.* **26**, 658–665 (1997).

4. Informationsberichte des Bayerischen Landesamtes für Wasserwirtschaft 1/97 *Grundwasserversauerung in Bayern* (Eigenverlag, München, 1997).

5. Armbruster, M. *Bodenkundliche Abhandlungen* **38**, 1–301 (1998).

6. Vesely, J., Hruska, J., Norton, S. A. & Johnson, C. E. *Maj. Solutes* **108**, 107–127 (1998).

7. Lischeid, G. *J. Hydrol.* (submitted).

8. Dise, N. *et al.* in *Acidification of Freshwater Ecosystems: Implications for the Future* (eds Steinberg, C. E. W. & Wright, R. F.) (Wiley, Chichester, 1994).

9. Likens, G. E., Driscoll, C. T. & Buso, D. C. *Science* **272**, 244–246 (1996).

Stoddard *et al.* reply — Alewell *et al.* introduce some important factors influencing the process of ecosystem recovery from acidification. But whereas we exclusively considered aquatic acidification and recovery¹, they focus on soil acidification.

First, Alewell *et al.* suggest that regions with soils that strongly adsorb sulphate (SO_4^{2-}) may show delayed recovery because of desorption of sulphate and the associated mobilization of base cations. Indeed, sulphate desorption is likely to occur over long periods and will delay recovery, particularly in regions typified by sulphur-adsorbing soils. Such soils are common in parts of Germany, as well as in the southeastern United States². The primary effect of desorption of soil-stored sulphate will be smaller decreases in surface-water sulphate over time. This effect is factored into our analyses, because we focused on the surface-water response¹. If the SO_4^{2-} trends we report are smaller than those predicted from declines in deposition, they still represent the net effect of changes in deposition and soil storage.

From the perspective of the large-scale responses we attempted to characterize, one might question whether sulphate-adsorbing soils are common enough to dominate a regional response. Of the 15 German sites we studied, four are in regions with sulphate-adsorbing soils (although three of the four show strong downward trends in surface-water SO_4^{2-}). For a regional trend to be valid, it must be homogeneous across the region (we tested this statistically¹) and the sites must be representative of the region. If

four sites are not enough to represent the sulphur-sorbing portion of the region, then our assumption that “national monitoring programmes have made good judgements in choosing sites representative of acid-sensitive surface waters in the region”¹ may be wrong. But we know of no data (for example, from a probability survey³ of German surface waters) to support this.

Second, Alewell *et al.* assert that increases in surface water alkalinity do not reflect recovery from soil acidification, and that the leaching of soil cations that occurs as sulphur desorbs from soils may result both in increased surface-water alkalinity and depleted soil cations. If one defines recovery in terms of restoring soil base saturation, none of the regions for which we report recovery of surface waters is likely to be experiencing soil recovery — all are continuing to receive large inputs of acidic deposition, all are exporting base cations from their watersheds at increased rates, and all will probably require long periods (centuries) of decreased deposition before weathering of parent material leads to widespread improvements in soil base saturation. The focus of our analysis¹ was on surface-water acidification and recovery, not on ‘ecosystem’ recovery — recovery of soils also needs to be a goal of emission-reduction programmes.

Third, Alewell *et al.* claim that North American regions should show greater recovery because deposition has been declining for longer in North America than in Europe. We argued that northern Europe has been experiencing much larger deposition declines¹ and for longer than North America^{4,5}, and that some of the failure of recovery we observe in North American regions may simply be a lagged response.

Deposition in the northeastern United States began to decline in the early 1970s (ref. 4), whereas deposition in the Nordic countries was already declining in the late 1960s (ref. 5); we also note that declines in percentage deposition have been larger in Europe¹. But the relative percentages in sulphate decline¹ tell only part of the story — in Europe between 1985 and 1996, the decline was from about 2.4 mg SO_4^{2-} per litre to 1.5 mg SO_4^{2-} per litre (a fall of 0.9 mg l^{-1}), whereas in North America the decline over the same time period was from 1.05 mg SO_4^{2-} per litre to 0.75 mg SO_4^{2-} per litre (a fall of 0.3 mg l^{-1}). Thus, if deposition declines are driving the changes in surface-water chemistry we report¹, then these have been more forceful in Europe than in North America in recent decades.

J. L. Stoddard*, **D. S. Jeffries†**, **A. Lükewille‡**,
M. Forsius§, **J. Mannio§**, **A. Wilander||**

*Environmental Protection Agency, 200 SW 35th Street, Corvallis, Oregon 97333, USA

†Environment Canada, PO Box 5050, Burlington, Ontario L7R 4A6, Canada

‡Norwegian Institute for Air Research, PO Box 100,
2027 Kjeller, Norway

§Finnish Environment Institute, Box 140,
00251 Helsinki, Finland

||University of Agricultural Sciences, PO Box 7050,
750 07 Uppsala, Sweden

1. Stoddard, J. L. *et al.* *Nature* **401**, 575–578 (1999).
2. Church, M. R., Shaffer, P. W., Eshleman, K. N. & Rochelle, B. P. *Wat. Air Soil Pollut.* **50**, 39–48 (1990).
3. Stoddard, J. L., Driscoll, C. T., Kahl, S. & Kellogg, J. *Ecol. Appl.* **8**, 288–299 (1998).
4. Husar, R. B., Sullivan, T. J. & Charles, D. F. in *Acidic Deposition and Aquatic Ecosystems: Regional Case Studies* (ed. Charles, D. F.) 65–82 (Springer, New York, 1991).
5. Mylonas, S. *Trends of Sulphur Dioxide Emissions, Air Concentrations and Depositions of Sulphur in Europe since 1880* (International Cooperative Programme for the Monitoring and Evaluation of Long Range Transmission of Air Pollutants in Europe/Meteorological Synthesizing Centre-West, 1994).

Palaeoecology

A 3,000-year record of penguin populations

There are no historical records of changing penguin populations in the maritime Antarctic^{1,2}. Here we analyse^{3,4} the concentration of 'bio-elements' in a lake-sediment core dating back approximately 3,000 radiocarbon years. We found that the deposition of penguin droppings had a significant effect on the geochemical composition of the sediment core. Changes in sediment geochemistry reflect fluctuations in penguin numbers and suggest that variations in climate had an impact on penguin populations, which peaked somewhere between 1,400 and 1,800 years ago.

We collected sediment cores from a lake on the Ardley Peninsula (maritime Antarctica) during the fifteenth China Antarctic Research Expedition (December 1998–March 1999) using a 12-cm-diameter PVC pipe. One 67.5-cm core (Y2) spanned almost 3,000 radiocarbon years. This core was sectioned at 1.0-cm intervals for the upper 64 cm, with the bottom section being a consolidation of 64–67.5 cm. We collected, stored and processed samples using clean techniques, and then analysed the concentration of geochemical elements and isotope ¹³C.

We found that the concentration of elements such as sulphur, phosphorus (represented by P₂O₅), calcium (represented by CaO), copper, zinc, selenium, strontium, barium and fluorine in Y2 was much higher than in other lake sediments in the maritime Antarctic. The relative concentration of these elements was similar at different depths in the core (Fig. 1a). The average concentration of fluoride reaches about 7,700 p.p.m., almost 30 times the level in the soil of this region.

These unusually high levels of bio-elements result from penguin droppings, which are rich in these elements^{5,6}. The

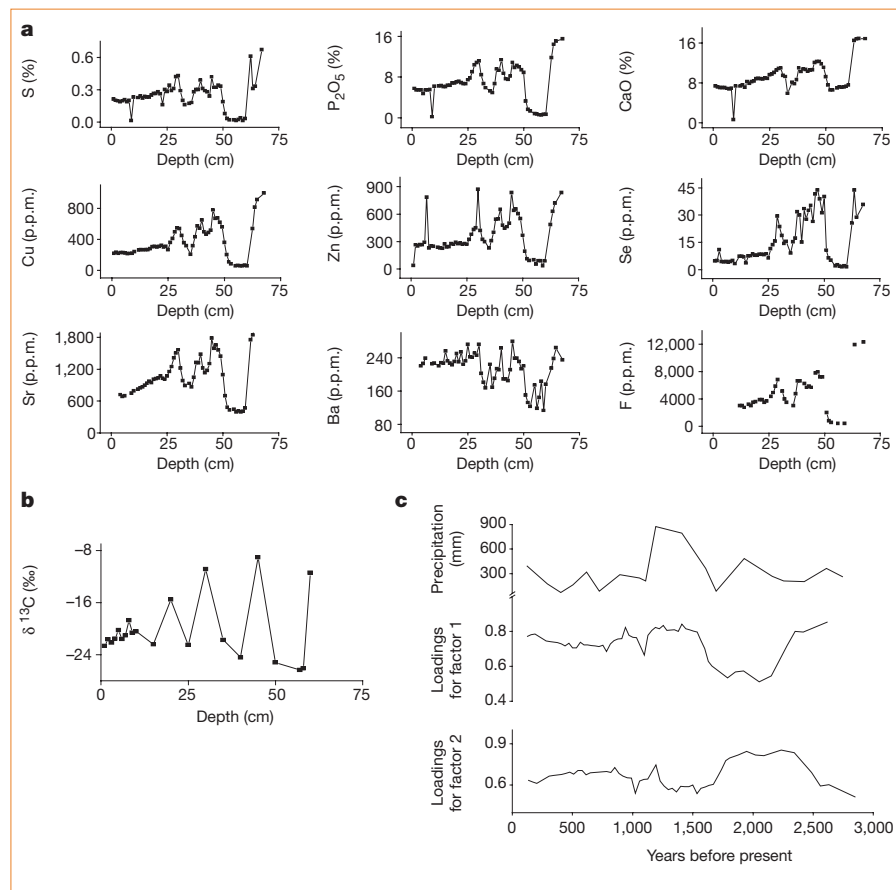


Figure 1 An isotope record of droppings from penguin populations. **a**, Concentration of elements S, P (as P₂O₅), Ca (as CaO), Cu, Zn, Se, Sr, Ba and F in Y2 lake sediment versus depth. **b**, Values of stable isotope δ¹³C of organic carbon in Y2 lake sediment. **c**, Comparison of the results from Q-mode analysis with precipitation; variation in loadings of factor 1 (middle curve) shows the variation in penguin populations during the past 3,000 years.

organic-carbon stable isotope (δ¹³C) values of between −9.038‰ and −26.327‰ (mean, −20.420‰) in Y2 lake sediment also indicate a contribution from penguin excreta^{7,8} (Fig. 1b). There is a penguin rookery around the lake, and their droppings are probably deposited in it both by catchment snow-melt and directly into the water.

We used Q-mode factor analysis^{3,4}, a method for decomposing multiple factors, to determine the concentration of bio-elements in the sediments. This revealed that 99.8% of the variance in the data could be explained by two factors (54.7% for factor 1 and 45.1% for factor 2). The first factor was defined to reflect changes in the input of penguin droppings to the sediment; the second factor is opposite to the first factor.

Deposition of seal hairs has been used as an index of changes in Antarctic mammal populations⁹. We have used changes in the deposition of bio-elements from penguin droppings as an indirect measure of population change. Thus, the first factor in Fig. 1c represents penguin population changes during the past 3,000 years. Our results show that the penguin population began to decline at 3,000 yr before present (BP) and was lowest at 1,800–2,300 yr BP, a

period of low temperature¹⁰. After this, the population increased, peaking between 1,400 and 1,800 yr BP. The peak corresponds almost exactly to a period of high precipitation¹¹ (Fig. 1c). Incremental sediment erosion cannot account for the high concentration of penguin droppings in the sediment. Historically, the size of penguin populations seems to be related to climate, and climate change might still affect the survival and abundance of modern penguin populations.

Liguang Sun, Zhouqing Xie, Junlin Zhao

Institute of Polar Environment, University of Science and Technology of China, Hefei, Anhui 230026, PR China

e-mail: slg@ustc.edu.cn

1. Taylor, R. H., Wilson, P. R. & Thomas, B. W. *Polar Rec.* **26**, 293–304 (1990).
2. Zhang, Z. W. & Zheng, G. M. *Chin. J. Zool.* **30**, 13–15 (1995).
3. Xu, Z. B. & Lou, Y. R. *Elements of Mathematical Geology* 226–238 (Beijing Univ. Press, 1994; in Chinese).
4. Wang, G. *Acta Geograph. Sinica* **53**, 538–545 (1998).
5. Tedrow, J. C. F. & Ugolini, F. C. *Am. Geophys. Union Antarct. Res. Ser.* **8**, 161–177 (1966).
6. Ugolini, F. C. & Starkey, R. L. *Nature* **211**, 440–441 (1966).
7. Smith, B. N. & Epstein, S. *Plant Physiol.* **47**, 380–384 (1971).
8. Stuiver, M. *Quat. Res.* **5**, 251–262 (1975).
9. Hodgson, D. A. & Johnston, N. M. *Nature* **387**, 30–31 (1997).
10. Clapperton, G. M. *et al. Quat. Res.* **31**, 210–228 (1989).
11. Zhao, J. L. *The Characteristics of the Modern Environmental Evolution in the Region of Antarctic Great Wall Station* (Science Press, Beijing, 1991).

Glacial/interglacial variations in atmospheric carbon dioxide

Daniel M. Sigman* & Edward A. Boyle†

* Department of Geosciences, Guyot Hall, Princeton University, Princeton, New Jersey 08544, USA

† Department of Earth, Atmospheric, and Planetary Sciences, Room E34-258, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Massachusetts 02139, USA

Twenty years ago, measurements on ice cores showed that the concentration of carbon dioxide in the atmosphere was lower during ice ages than it is today. As yet, there is no broadly accepted explanation for this difference. Current investigations focus on the ocean's 'biological pump', the sequestration of carbon in the ocean interior by the rain of organic carbon out of the surface ocean, and its effect on the burial of calcium carbonate in marine sediments. Some researchers surmise that the whole-ocean reservoir of algal nutrients was larger during glacial times, strengthening the biological pump at low latitudes, where these nutrients are currently limiting. Others propose that the biological pump was more efficient during glacial times because of more complete utilization of nutrients at high latitudes, where much of the nutrient supply currently goes unused. We present a version of the latter hypothesis that focuses on the open ocean surrounding Antarctica, involving both the biology and physics of that region.

The past two million years have been characterized by large cyclic variations in climate and glaciation. During cold 'ice age' periods, large continental ice sheets cover much of the polar Northern Hemisphere. During intervening warm periods, or 'interglacials', Northern Hemisphere glaciation wanes drastically. The ultimate pacing of these glacial cycles is statistically linked to cyclic changes in the orbital parameters of the Earth, with characteristic frequencies of roughly 100, 41 and 23 kyr (refs 1, 2). These orbitally driven variations in the seasonal and spatial distribution of solar radiation incident on the Earth's surface, known as the 'Milankovitch cycles' after their discoverer, are thought to be the fundamental drivers of glacial/interglacial oscillations. However, the direct energy budget effects of the orbital variations are insufficient to drive the large amplitude of the glacial cycles that are observed, and orbital variations alone do not provide an obvious cause of the rapid climate transitions evident in palaeoclimatic and palaeoceanographic records. Positive feedbacks within the Earth's climate system must amplify orbital forcing to produce glacial cycles, but the operation of these internal feedbacks is poorly understood. One of the central goals of Earth science is to develop a mechanistic understanding of the Earth's climate feedbacks, the roles that they have played over glacial cycles, and the roles that we should expect them to play in the future.

The concentration of carbon dioxide (CO_2) in the atmosphere has varied in step with glacial/interglacial cycles^{3,4} (Fig. 1). During interglacial times, such as the Holocene (roughly the past 10,000 years), the atmospheric partial pressure of CO_2 (p_{CO_2}) is typically near 280 parts per million by volume (p.p.m.v.). During peak glacial times, such as the Last Glacial Maximum about 18,000 years ago, atmospheric p_{CO_2} is 180–200 p.p.m.v., or roughly 80–100 p.p.m.v. lower. CO_2 is a greenhouse gas, and model calculations suggest that its changes play a significant role in the energetics of glacial/interglacial climate change⁵. However, we have not yet identified the cause of these variations in CO_2 . How do the orbitally driven cycles in solar radiation set this particular positive feedback into motion?

This question has persisted for two decades, motivating intensive research by palaeoclimatologists and palaeoceanographers. From this research, we now recognize many processes that could regulate atmospheric CO_2 on the timescales of glacial/interglacial transitions, and it may well be that the actual cause of glacial/interglacial CO_2 change is among them. However, evaluating the importance

of these mechanisms with data from the recent geological and glaciological record has been a challenging and controversial task, leading as yet to no consensus on a fundamental mechanism. It is possible that many factors contribute comparably to the observed changes. However, the regularity of the CO_2 variations and the consistency of the upper and lower limits of atmospheric CO_2 through multiple 100-kyr cycles (Fig. 1) are suggestive of a well-ordered set of dominant mechanisms, the 'holy grail' of glacial/interglacial CO_2 research.

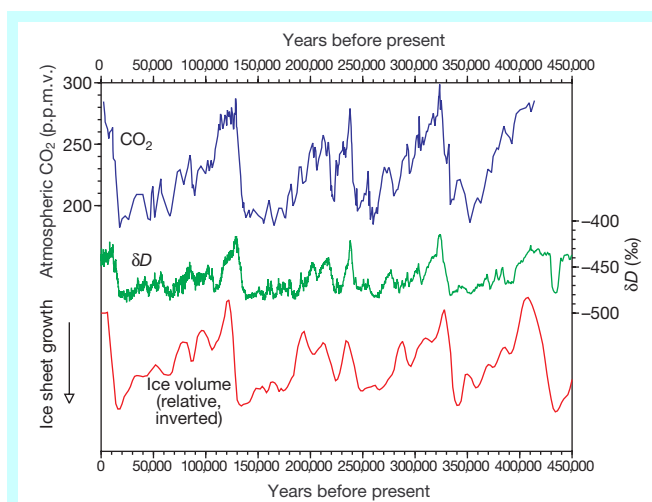


Figure 1 The history of atmospheric CO_2 back to 420 kyr ago as recorded by the gas content in the Vostok ice core from Antarctica⁴. The ratio of deuterium to hydrogen in ice (expressed as the term δD) provides a record of air temperature over Antarctica, with more negative δD values corresponding to colder conditions. The history of global ice volume based on benthic foraminiferal oxygen isotope data from deep-sea sediment cores⁹⁶ is plotted as relative sea level, so that ice ages (peaks in continental ice volume) appear as sea level minima, with a full glacial/interglacial amplitude for sea level change of about 120 m (ref. 18). During peak glacial periods, atmospheric CO_2 is 80–100 p.p.m.v. lower than during peak interglacial periods, with upper and lower limits that are reproduced in each of the 100-kyr cycles. Ice core records, including the Vostok record shown here, indicate that atmospheric CO_2 was among the early parameters to change at the termination of glacial maxima, roughly in step with Southern Hemisphere warming and preceding the decline in Northern Hemisphere ice volume.

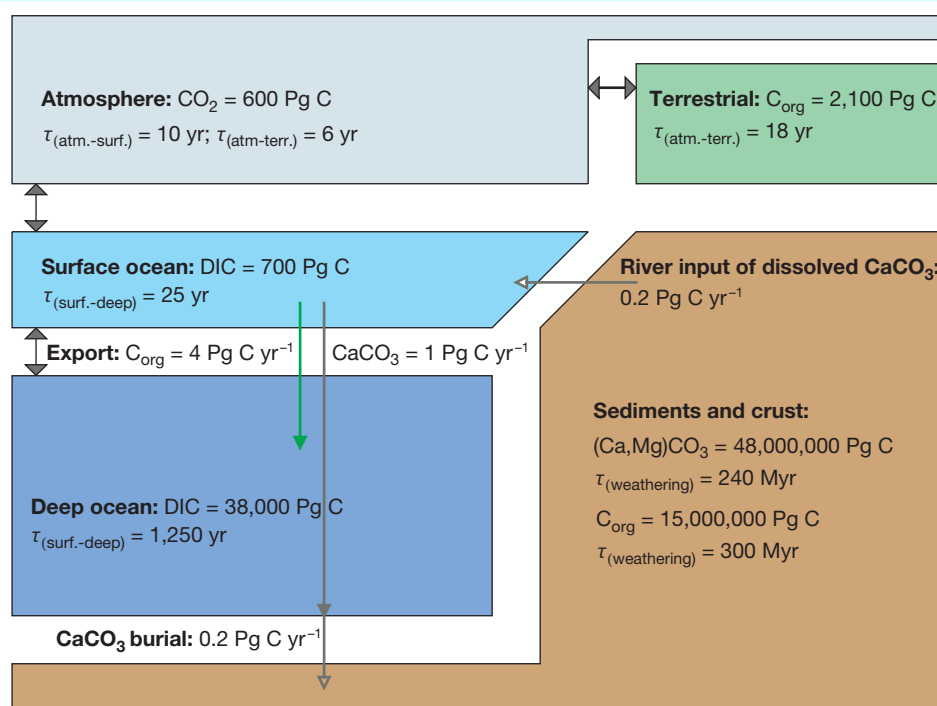


Figure 2 A simplified view of the Holocene (pre-industrial) carbon cycle. Data is largely based on ref. 97 and includes only the reservoirs and fluxes thought to be most relevant for glacial/interglacial variations in atmospheric CO_2 . Carbon that exchanges with atmospheric CO_2 on the timescale of glacial/interglacial transitions is stored in a number of reservoirs (given in units of Pg C, or 10^{15} g C), as dissolved inorganic carbon (DIC), calcium carbonate (CaCO_3), and organic carbon (C_{org}). For each of the carbon reservoirs, a residence time (τ , reservoir size divided by input or output) is given in relation to important exchange terms among reservoirs, which are indicated by double arrows. Relevant fluxes of C_{org} and CaCO_3 (including 'dissolved' CaCO_3) are indicated by single arrows. It is also informative to consider the fluxes that are excluded from this diagram. For instance, organic carbon burial (which occurs mostly along ocean margins), organic carbon weathering on land and volcanic/hydrothermal CO_2 release are generally thought to occur too slowly or to have the wrong sensitivities to play a major role in the consistent variation of atmospheric CO_2 over glacial/interglacial cycles.

The modern carbon cycle provides our first basis for discerning the processes that may have played an important role in glacial/interglacial CO_2 change (Fig. 2). Excluding recent anthropogenic changes, the atmospheric CO_2 reservoir is one-third to one-quarter of the size of the terrestrial biosphere and soil carbon reservoir, and is comparable to the amount of inorganic carbon stored in the warm upper ocean. However, all of these pools together are less than 10% of the inorganic carbon stored in the deep ocean. Deep water is exposed at the ocean surface roughly every 1,000 years, so that a change in atmospheric CO_2 driven by something other than oceanic processes would be diluted into the large reservoir of deep ocean inorganic carbon, greatly attenuating such CO_2 changes on the timescale of ice ages and interglacial periods. In addition, over thousands of years, the deep ocean regulates the exchange of inorganic carbon between the geosphere and the hydrosphere/atmosphere system through its control of the burial of calcium carbonate on the sea floor. From these facts, Broecker⁶ concluded that oceanic changes must underlie the glacial/interglacial cycle in atmospheric CO_2 .

Here, we review the potential causes of glacial/interglacial CO_2 change that have been recognized as well as the evidence for and against them. We will start with the relatively well-studied hypotheses and end with subjects of greater conjecture. Given that we still lack a proven mechanism for glacial/interglacial CO_2 changes, this approach is tantamount to first describing the processes that we believe cannot explain glacial/interglacial CO_2 changes (changes in terrestrial carbon storage and in ocean temperature, simple changes in the marine calcium carbonate budget) and then shifting to those processes that might (such as changes in the oceanic nutrient reservoir size) and finally to a hypothesis that is currently receiving special attention (changes in the completeness of nutrient consumption in high-latitude surface waters).

Carbon storage on land

With currently forested or vegetated regions covered under kilometres of ice, an increase in the extent of deserts, apparent conversion of tropical forest to grassland, and exposure of organic-rich sediments on continental shelves, it seems likely that the continental reservoir of organic carbon decreased during the last ice age, contributing CO_2 to the ocean/atmosphere system (ref. 7 and references therein). Based on reconstructions of biome distributions for the glacial world and information on carbon storage by different biomes, estimates^{7,8} of the glacial/interglacial change in terrestrial carbon storage range between 700 and 1,350 Pg C (10^{15} g of carbon). An integrative indicator of the transfer of carbon from terrestrial sources during the last ice age is thought to be provided by the $^{13}\text{C}/^{12}\text{C}$ ratio of the calcitic shells of benthic foraminifera from deep sea sediments. The $^{13}\text{C}/^{12}\text{C}$ ratio of glacial-age shells is on average 0.3–0.4‰ lower than that of Holocene shells, a shift that has been interpreted as reflecting a transfer of continental organic carbon, which has a low $^{13}\text{C}/^{12}\text{C}$ ratio, to the ocean/atmosphere inorganic carbon reservoir during ice ages^{9–11}. If this interpretation is correct, the $^{13}\text{C}/^{12}\text{C}$ change of 0.3–0.4‰ indicates that roughly 300–700 Pg of additional carbon were distributed throughout the ocean/atmosphere system during the last ice age¹². While other possible explanations¹³ have been raised for the observed decrease in foraminiferal $^{13}\text{C}/^{12}\text{C}$, the foraminiferal estimate of terrestrial carbon loss during the last ice age is already at the lower bound of values inferred from the terrestrial palaeoecological data, making it seem likely that changes in the terrestrial carbon pool dominated the benthic foraminiferal $^{13}\text{C}/^{12}\text{C}$ shift.

Given no other processes, a 500 Pg C addition, when distributed throughout the ocean/atmosphere system, would have raised atmospheric CO_2 by 45 p.p.m.v. However, this increase in inorganic

carbon would have driven a temporary increase in the dissolution of deep sea sedimentary calcite (see below), which would have reduced the effective CO_2 increase to 15 p.p.m.v. over about 5–10 kyr. Thus, while the terrestrial carbon reservoir was smaller during the last ice age, its effect on atmospheric CO_2 was probably modest because of buffering by the ocean carbon reservoir and the ocean CaCO_3 balance. Details aside, terrestrial carbon represents a source of CO_2 to the atmosphere, not a sink, during ice ages.

CO_2 storage in the ocean

The exchange of CO_2 between the atmosphere and the surface ocean would reach completion over the course of six to twelve months if there were no other processes redistributing inorganic carbon in the ocean. However, the p_{CO_2} of surface waters is continuously being reset by its interaction with the deep ocean reservoir of inorganic carbon, which is more than 25 times that of the atmosphere and surface ocean combined (Fig. 2).

Excluding the effect of anthropogenic additions of CO_2 to the atmosphere, we can consider the ocean and atmosphere to be in approximate steady state with respect to CO_2 exchange, with CO_2 uptake in some regions of the ocean balanced by CO_2 release in others. The air–sea exchange of CO_2 is determined largely by the p_{CO_2} of surface waters, with CO_2 transfer into the ocean occurring where the p_{CO_2} of surface waters is lower than that of the atmosphere, and transfer out of the ocean where it is higher. Surface water p_{CO_2} , in turn, is controlled by temperature and ocean chemical composition, which are separable in their responses to glacial/interglacial changes. CO_2 is more soluble in cold water, causing CO_2 invasion into cold surface waters and CO_2 evasion out of warm surface waters. The role of ocean chemical composition is more complicated, involving the dependence of surface water p_{CO_2} on its inorganic carbon content and acid/base chemistry, which are greatly affected by biological processes.

Ocean temperature

As ocean temperature was lower during ice ages, it is an obvious first step to consider its effect on atmospheric CO_2 . The lower temperatures of the glacial ocean would have reduced the concentration of CO_2 in the atmosphere by drawing more of it into the ocean. The deep ocean, which is the dominant volume of ocean water, has a mean temperature of 2°C . Sea water begins to freeze at about -2°C , producing buoyant ice. As a result, deep ocean water could not have been more than 4°C colder during the last ice age, placing an upper bound on how much additional CO_2 this water could have sequestered simply by cooling. The potential cooling of surface waters in polar regions such as the Antarctic is also constrained by the freezing point of sea water. The temperature of the lower-latitude (equatorial, tropical and subtropical) surface ocean is far from the freezing point and thus lacks this constraint, and the pattern of low-latitude cooling is a subject of continuing research. The CLIMAP project¹⁴ estimated that the tropical and subtropical sea surface was cooler by 3°C or less during the last ice age. However, estimates from the continents and some of the more recent paleoceanographic work suggest greater cooling, for example, as much as 5°C in the tropical Atlantic¹⁵.

The cooling of different regions of the ocean surface affected ice age CO_2 levels to the degree that they altered the volume-averaged temperature of the ocean, the critical parameter for the mean solubility of CO_2 in the ocean. The similarity in temperature between the large volume of the deep ocean and polar surface waters implies that most of the ocean interior is ventilated in polar environments, so we expect a given temperature change in polar surface waters to have a greater influence on atmospheric CO_2 than would the same temperature change in the low-latitude surface. For instance, in the CYCLOPS ocean box model¹⁶, a given cooling of the high-latitude surface causes roughly twice the CO_2 decrease that is

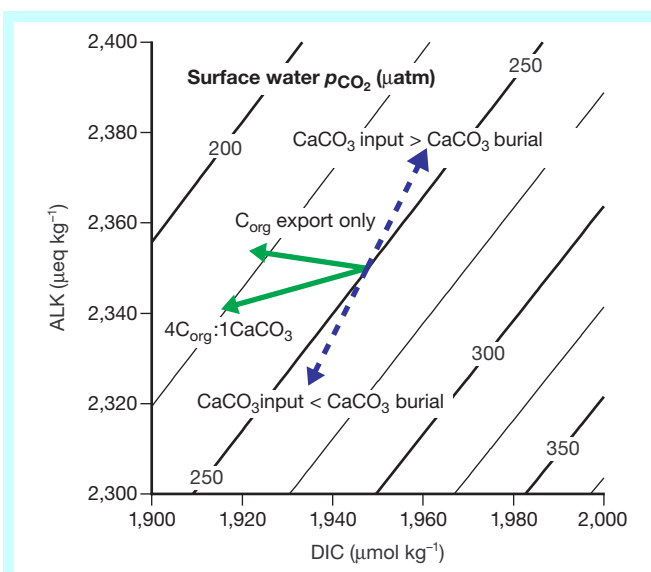


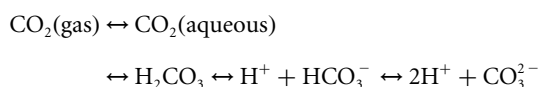
Figure 3 The p_{CO_2} of surface sea water at 20°C and a salinity of 35 parts per thousand, as set by its dissolved inorganic carbon (DIC) content and alkalinity (ALK). The p_{CO_2} decreases as DIC decreases and/or ALK increases. The dashed arrows show the relative effects on ALK and DIC of imbalances in the CaCO_3 budget of the ocean. If CaCO_3 burial is less than the continental input of dissolved CaCO_3 , then the ALK and DIC of the whole ocean increase in a 2:1 ratio. This lowers the p_{CO_2} of surface waters and transfers CO_2 from the atmosphere into the ocean. The solid arrows show the effect of the rain of biogenic materials out of the surface ocean ('export production'), for the cases that this rain is organic carbon (C_{org}) only and that it is a 4:1 molar ratio of C_{org} and CaCO_3 . In general, the former composition applies to high nutrient polar surface ocean ecosystems, while the latter applies to the warm low-latitude ocean (references in ref. 98). Export production removes DIC from the surface ocean, sequestering it in the ocean interior. The modest ALK increase associated with C_{org} export only is due to the associated uptake of nitrate by phytoplankton. Export production also affects the DIC and ALK of the deep ocean, because C_{org} produced in the surface ocean and oxidized by bacteria in the ocean interior releases DIC and lowers the carbonate ion concentration of deep water. This affects the burial rate of CaCO_3 in the seafloor sediments, which further modifies the p_{CO_2} of surface waters through its effect on the whole-ocean CaCO_3 balance (dashed arrows).

caused by an identical cooling of the low-latitude surface. A reasonable estimate for ice age ocean cooling of 2.5°C in the polar surface and 5°C in the low-latitude surface leads to a CO_2 decrease of 30 p.p.m.v., with the low-latitude and polar temperature changes playing roughly equivalent roles in this decrease. It has recently been noted that ocean general circulation models predict a greater sensitivity of CO_2 to low-latitude surface conditions than do simple ocean box models like CYCLOPS¹⁷. However, the significance of this observation is a matter of intense debate.

An opposing effect on atmospheric CO_2 to that of glacial/interglacial temperature change is provided by the increased salinity of the glacial ocean, due to the storage of fresh water on land in extensive Northern Hemisphere ice sheets. Based on the approximately 120 m depression of sea level during the last ice age¹⁸, the whole ocean was about 3% saltier than it is today. All else being constant, this increase would have reduced the solubility of CO_2 in sea water and raised atmospheric CO_2 by 6.5 p.p.m.v. Taking the estimated temperature and salinity effects together, we would expect the atmospheric CO_2 concentration of the ice age world to have been 23.5 p.p.m.v. lower. Folding in the effect of a 500 Pg C transfer from the continents to the ocean/atmosphere system, we are left with an 8.5 p.p.m.v. decrease in CO_2 (Table 1). There are uncertainties in each of these effects, but it seems that most of the 80–100 p.p.m.v. CO_2 change across the last glacial/interglacial transition must be explained by other processes. We must move on to the more complex aspects of the ocean carbon cycle.

Ocean carbon cycle

Upon dissolution in sea water, inorganic carbon maintains an acid/base equilibrium:



Dissolved inorganic carbon (DIC) is defined as the sum of these different inorganic carbon species, excluding gas phase CO_2 . The distribution of inorganic carbon among the above species depends on both the DIC concentration and the acid-titrating capacity of sea water, which we refer to as 'alkalinity' (ALK), the total equivalence of ions that must be protonated in order to lower the pH of sea water to the pK_a of carbonic acid (Fig. 3). If DIC increases while ALK remains constant, the proportion of the DIC pool present as carbonic acid and aqueous CO_2 increases. In addition, there is more inorganic carbon to partition among the different inorganic carbon species. Hence, pCO_2 increases as DIC increases. As ALK increases, the inorganic carbon reservoir is shifted away from aqueous CO_2 , resulting in a lower pCO_2 .

The DIC and ALK of surface water are determined by two types of carbon cycle responses to oceanographic change, which can be referred to as 'closed-system' and 'open-system' responses. Closed-system responses redistribute DIC and ALK among the different volumes of the ocean. Open-system responses change the sizes of the whole-ocean reservoirs of DIC and ALK, through temporary imbalances between their input from weathering of calcium carbonate (CaCO_3) on the continents and their output by the burial of biogenic calcium carbonate in the ocean. The first geochemical hypotheses for the cause of glacial/interglacial CO_2 changes generally fell into two categories, based on whether they elicited a closed-system or open-system carbon-cycle response. Closed-system hypotheses focused on the 'biological pump', the sequestration of inorganic carbon in the deep ocean by photosynthesis in the sunlit surface and the subsequent rain of organic carbon out of the upper ocean¹⁹. Open-system hypotheses focussed on carbonate weathering and coral reef growth²⁰. We have come to recognize that many geochemical changes, especially those involving the biological pump, would affect both the gradients and the reservoir sizes of ALK and DIC in the ocean, so that both closed- and open-system responses must be considered²¹.

Ocean alkalinity

Alkalinity is added to the ocean largely by the weathering of emergent carbonates on land, and it is removed largely by the biogenic precipitation and burial of CaCO_3 , in shallow coral reefs and carbonate banks, and in deep-sea sediments²². A complete treatment of ocean alkalinity would also include the input of alkalinity by the weathering of silicates on land and the consumption of alkalinity in the ocean by hydrothermal activity²³ and 'reverse weathering'²⁴. However, on the timescales of glacial/interglacial transitions, it is believed that our view can be restricted to the input/output balance of CaCO_3 . In reef environments, much of the buried CaCO_3 is precipitated *in situ* by corals and benthic algae as the mineral aragonite. In the open ocean, most CaCO_3 is precipitated as the mineral calcite in the microfossils of two groups of marine plankton, coccolithophorids and planktonic foraminifera. This CaCO_3 subsequently rains out of the surface ocean along with

sinking organic matter. Roughly 75% of the CaCO_3 rain dissolves in the deep sea, while the remaining 25% is buried in deep sea sediment and is thus removed from the ocean reservoir (Fig. 2). When weathered or buried, CaCO_3 adds or removes two equivalents of ALK per mole of DIC.

The balance between the input of alkalinity from the continents and the removal of alkalinity in sediments is maintained by the distribution of CaCO_3 burial in the deep sea, and this distribution also provides us with a powerful constraint on ocean alkalinity changes over glacial/interglacial cycles. Deep sea CaCO_3 burial is mediated by the depth of the 'calcite lysocline,' the transition separating shallower seafloor sediments where some calcite is preserved from deeper seafloor sediments where almost all of the calcite rain dissolves^{25,26}. One of the primary controls on the depth of the lysocline is the calcite saturation depth, the depth at which the carbonate ion concentration of bottom water, $[\text{CO}_3^{2-}]_{\text{bw}}$, equals the carbonate ion concentration at calcite saturation, $[\text{CO}_3^{2-}]_{\text{sat}}$, occurring on average at 3.5 km depth²⁷. Because calcite solubility increases with depth (a pressure effect), the water bathing the sea floor above the saturation depth is supersaturated, causing the preservation and eventual burial of CaCO_3 on the shallow sea floor, while water deeper than the saturation depth is undersaturated, dissolving the calcite rain and thus preventing CaCO_3 burial on the deeper sea floor.

Because of its effect on the calcite burial rate, the response of the lysocline to a change in $[\text{CO}_3^{2-}]_{\text{bw}}$ represents a negative feedback which drives deep-sea calcite burial towards a balance with the input of alkalinity from the continents²⁸. For example, an increase in the weathering of limestone and other carbonate-bearing rocks on land (or similarly a decrease in coral reef growth in the shallow ocean) increases the whole ocean inventories of ALK and DIC in a 2:1 ratio. This inventory change causes a decrease in atmospheric CO_2 (Fig. 3). However, this increase in oceanic alkalinity also causes an increase in deep ocean $[\text{CO}_3^{2-}]$ (according to the relation $\Delta[\text{CO}_3^{2-}] \approx 0.6\Delta(\text{ALK} - \text{DIC})$), which in turn deepens the calcite saturation horizon. This deepening of the saturation horizon deepens the lysocline, which increases the area of sea floor where CaCO_3 is buried, so that the CaCO_3 burial rate once again comes into balance with the rate of alkalinity input from rivers, and atmospheric pCO_2 stabilizes at a lower value.

Thus, the history of the mean lysocline depth provides a palaeoceanographic constraint on ocean alkalinity changes. An increase in carbonate weathering on land would deepen the lysocline by 1 km for every ~ 25 p.p.m. decrease in atmospheric pCO_2 , that it would cause. This requires that the lysocline must have been deeper by more than 2 km during the last ice age for a change in the input of alkalinity to have played a dominant role in glacial/interglacial CO_2 changes. Data on seafloor CaCO_3 content suggest that the ocean-average lysocline of the last ice age was less than 1 km deeper than that of the Holocene²⁹. This finding invalidates hypotheses that call upon increased weathering or decreased coral reef growth as the primary driver of lower ice age levels of atmospheric CO_2 (ref. 30).

An alternative mechanism for changes in ocean alkalinity involves the biological production of CaCO_3 microfossils. The low-latitude ocean (the combined equatorial, tropical, and subtropical regions), where the $\text{CaCO}_3/\text{C}_{\text{org}}$ ratio of the biogenic rain is roughly 1:4, is responsible for most deep-sea calcite burial²⁹. We might expect the $\text{CaCO}_3/\text{C}_{\text{org}}$ rain ratio of the low-latitude ocean to have been lower during the last ice age if organic carbon export production rates were higher or more sporadic, amongst other possibilities³¹. Decreasing the $\text{CaCO}_3/\text{C}_{\text{org}}$ rain ratio of the low-latitude ocean drives a decrease in atmospheric CO_2 , partly by leaving a larger fraction of the ocean's alkalinity reservoir in the surface ocean, where it enhances CO_2 solubility (a closed-system effect)³¹, and partly by decreasing deep-sea CaCO_3 burial, which causes the whole ocean alkalinity reservoir to increase (an open-system effect)³². In the CYCLOPS ocean geochemistry model¹⁶, assuming that only

Table 1 Atmospheric CO_2 effects of known changes

Condition during the last ice age (as different from Holocene)	CO_2 change (p.p.m.v.)
Terrestrial carbon decrease (500 Pg C)	15
Ocean cooling (5° low latitude, 2.5° high latitude)	-30
Ocean salinity increase (3%)	6.5
Total CO_2 change	-8.5

bottom water undersaturation causes calcite dissolution on the sea floor²⁷, a halving of the $\text{CaCO}_3/\text{C}_{\text{org}}$ rain ratio causes atmospheric CO_2 to decrease by about 55 p.p.m.v., with 35 p.p.m. of this change arising from an increase in whole ocean alkalinity (Fig. 4)³³. This whole ocean alkalinity change can be parsed into two separate effects. The first is 'CaCO₃ compensation'²¹, in which the closed-system redistribution of ALK and DIC by the change in rain ratio causes a 'dissolution event' in the deep sea, reducing the depth of the lysocline (and thus the whole-ocean burial rate of CaCO₃) for a transient period (Fig. 4). The second open-system effect, which we call the 'steady-state lysocline shift'³³, is that the decrease in the rain rate of CaCO₃ to the sea floor in itself causes the burial rate of CaCO₃ to fall below the continental input rate of dissolved CaCO₃. This, in turn, causes ocean alkalinity to increase until the lysocline has deepened enough to allow for CaCO₃ burial to once again balance the continental input of alkalinity.

The 55 p.p.m.v. decrease in CO_2 due to this rain ratio change is significant relative to the observed glacial/interglacial variability of atmospheric CO_2 . However, the record of deep sea calcite preservation provides a critical sticking point, just as it did with respect to an increase in the continental sources of alkalinity. The CO_2 decrease is associated with an ocean-average 1.1-km deepening of the calcite saturation horizon (Fig. 4) owing to the decrease in the

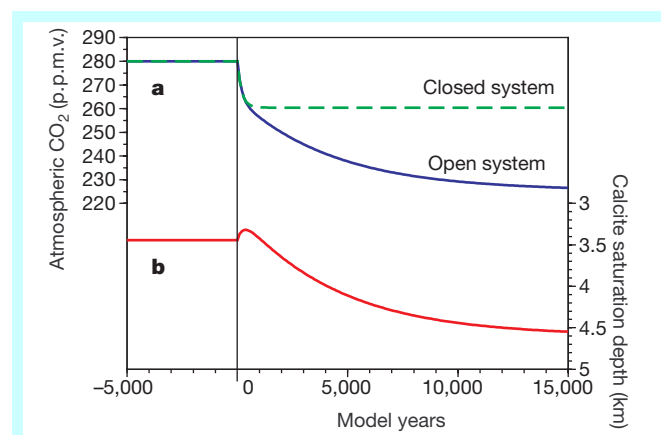


Figure 4 The time-dependent response of the CYCLOPS model to a sudden halving in the $\text{CaCO}_3/\text{C}_{\text{org}}$ ratio of the export flux out of the low-latitude surface ocean³³. The low-latitude surface ocean comprises the equatorial, tropical and subtropical oceans. **a**, Atmospheric CO_2 , for both 'closed' and 'open' system cases. **b**, Depth of calcite saturation (and the lysocline) in the open-system case. In the closed-system case, all of the CaCO_3 rain is dissolved in the ocean subsurface, with no burial, and there is no continental source of dissolved CaCO_3 . In the open-system case, which is more realistic, CaCO_3 is buried on the deep sea floor, and there is a continental input of dissolved CaCO_3 . The closed-system output has been offset to start at the same initial p_{CO_2} as the open-system case (280 p.p.m.v.). The closed-system effect of the $\text{CaCO}_3/\text{C}_{\text{org}}$ rain ratio decrease is to extract less alkalinity from the surface ocean, which causes a 20 p.p.m.v. decrease in atmospheric CO_2 within hundreds of years. In the open-system case, the $\text{CaCO}_3/\text{C}_{\text{org}}$ rain ratio decrease causes CaCO_3 burial to decrease below the input rate of alkalinity from the continents, both because of a transient dissolution event (evident as the early shoaling of the calcite saturation horizon, **b**) and because of a permanent decrease in the CaCO_3 rain rate onto the sea floor. Due to both of these effects, ocean alkalinity increases, which causes the p_{CO_2} of surface water to decrease, so that atmospheric CO_2 decreases more in the open-system case than in the closed-system case (**a**). The whole-ocean alkalinity increase associated with the permanent decrease in the CaCO_3 rain rate, while contributing to the decrease in atmospheric CO_2 , also causes the deep-ocean carbonate ion concentration to increase through time, so that the calcite saturation horizon deepens. This deepening increases the area of sea floor on which CaCO_3 is preserved and buried so that global-ocean CaCO_3 burial increases back to equality with the river input of alkalinity, resulting in a new steady state with a deeper saturation horizon and lysocline. The model used here assumes that the carbonate ion concentration of deep water is the dominant control on the depth distribution of calcium carbonate burial in seafloor sediments (see text).

CaCO_3 rain rate to the sea floor. The compiled data indicating a smaller difference in the ocean-average lysocline depth between the modern and glacial oceans²⁹ argue against a glacial decrease in the $\text{CaCO}_3/\text{C}_{\text{org}}$ ratio as the sole driver of lower glacial CO_2 levels. What is needed to explain glacial/interglacial CO_2 variability is a mechanism to lower atmospheric CO_2 without causing a large deepening of the steady-state lysocline.

Early research demonstrated that the degree of seawater saturation with respect to calcite is a dominant control on seafloor CaCO_3 preservation. The observed correspondence of the lysocline depth with the depth of calcite saturation helped lead to this conclusion²⁷. However, continued studies of seafloor dissolution have indicated that pore water undersaturation due to oxidation of organic matter within the sediments is also a significant driver of calcite dissolution (ref. 34 and references in ref. 35). The effect of sedimentary respiration on seafloor calcite dissolution amplifies the capacity of a decrease in the $\text{CaCO}_3/\text{C}_{\text{org}}$ ratio of low-latitude biological production to lower atmospheric CO_2 (ref. 30). When this 'respiratory dissolution' is included in an ocean geochemistry model, halving the $\text{CaCO}_3/\text{C}_{\text{org}}$ ratio causes an atmospheric CO_2 decrease comparable to the glacial/interglacial change. The greater CO_2 decrease is due to enhancement of the whole ocean alkalinity increase when respiratory dissolution is included in the seafloor dissolution model. One of the recognized requirements of this mechanism as a hypothesis for glacial/interglacial CO_2 changes is that the deepening of the calcite saturation horizon associated with this alkalinity increase would not be evident in the record of the glacial lysocline. In other words, it requires that the depth of lysocline did not correspond to the calcite saturation depth during glacial times. Instead, the lysocline must have occurred in supersaturated waters, representing the transition point at which bottom water supersaturation was low enough for respired CO_2 in sedimentary pore waters to drive almost complete dissolution of the calcite in the sediment.

When the $\text{CaCO}_3/\text{C}_{\text{org}}$ rain ratio is lowered in a range of models of seafloor calcite dissolution, however, respiratory dissolution does not effectively decouple the lysocline from the saturation depth until both are quite deep, so that a major deepening in the calcite saturation horizon would have been accompanied by a deepening of the lysocline³³. The strong coupling between the lysocline and the saturation horizon results from the weak depth dependence of respiratory dissolution, relative to that of dissolution driven by bottom water undersaturation. With enhanced respiratory dissolution, calcite burial rates become less depth-dependent, with proportionally more dissolution occurring on the shallow sea floor. With less CaCO_3 burial on the sea floor of the shallow and mid-depth ocean, there must be an increase in CaCO_3 burial rates in the deeper ocean, and thus a deepening of the lysocline, to balance the river input of alkalinity. Thus, the relatively constant ocean-average lysocline depth remains as evidence against a glacial decrease in the $\text{CaCO}_3/\text{C}_{\text{org}}$ ratio as the sole driver of lower ice age CO_2 levels. Although seafloor CaCO_3 dynamics may have surprises in store for us, there are reasons to doubt the role of whole-ocean alkalinity changes as the sole driver of glacial/interglacial CO_2 change.

Nevertheless, most biological pump mechanisms for lowering atmospheric CO_2 also drive a change in whole-ocean alkalinity (see below), and in some cases, this contributes significantly to their capacity to lower atmospheric CO_2 . The alkalinity changes which are most 'helpful' in this regard are due to CaCO_3 compensation²¹, in which the oceanic redistribution of ALK and DIC causes a transient dissolution event, rather than a shift in the steady-state depth of the lysocline. This prevents violation of the palaeoceanographic constraint on the steady-state depth of the lysocline.

The ocean nutrient reservoir and low-latitude productivity

Broecker¹⁹ first hypothesized a glacial increase in the strength of the biological pump as the driver of lower glacial CO_2 levels. This

hypothesis has spawned many variants. These mechanisms can be separated into two groups: (1) those involving changes in the low- and mid-latitude surface ocean, where the major nutrients nitrate and phosphate ultimately limit the extraction of CO₂ by biological production; and (2) those involving changes in the polar and subpolar ocean regions, where the major nutrients are currently not completely consumed. In either case, the relevant quantity is 'export production', the organic matter produced by phytoplankton that is exported from the surface ocean, resulting in the sequestration of its degradation products (inorganic carbon and nutrients) in the ocean interior.

Biological production in the ocean tends to incorporate phosphate, nitrate, and inorganic carbon into biomass in the ratios of 1:16:106 (the 'Redfield ratios'³⁶). These ratios determine the amount of inorganic carbon sequestered in the deep sea when the supply of nutrients to the surface is completely converted to export production, as is the case in the low- and mid-latitude ocean. For low-latitude carbon export to extract more CO₂ from the surface ocean and atmosphere, it must be driven by an increase in the deep-ocean nutrient reservoir and/or a change in the composition of the export production.

Broecker¹⁹ and McElroy³⁷ described possible mechanisms by which the reservoirs of phosphate and nitrate might increase during glacial times, which would allow enhanced low-latitude biological production to lower atmospheric CO₂. However, it was recognized that large changes in nutrient reservoirs would be required to produce the entire observed amplitude of CO₂ change. In the modern polar oceans, biological production is limited by parameters other than the supply of phosphate or nitrate³⁸. Therefore, it seems most reasonable to assume that high-latitude production would not respond to an increase in these nutrients. Making this assumption, box model calculations³³ indicate that the immediate effect of 30% increase in the oceanic nutrient reservoir, enhanced extraction of CO₂ from the surface ocean and its sequestration in the deep sea, lowers atmospheric CO₂ by 30–45 p.p.m.v. (Table 2).

The longer-term effect on CO₂ from the ocean alkalinity balance depends greatly on whether CaCO₃ export increases in step with C_{org} export³³. If CaCO₃ export increases proportionately with C_{org} export, the increase in the CaCO₃ flux to the sea floor causes a loss of alkalinity from the ocean over 5–10 kyr, which works to raise atmospheric CO₂. As a result, the net drop in CO₂ from a 30% increase in the oceanic major nutrient reservoir would be only 15–25 p.p.m.v. (Table 2)³³. However, if CaCO₃ production remains constant as C_{org} production increases, an increase in the oceanic nutrient reservoir is much more effective at lowering atmospheric CO₂. In this case, a 30% increase in the oceanic nutrient reservoir lowers atmospheric CO₂ by approximately 50 p.p.m.v. (Table 2), and an increase of more than 50% could lower atmospheric CO₂ by 80 p.p.m.v.³³. The increase in the ocean nutrient reservoir required for an 80 p.p.m.v. decrease in CO₂ could be lowered from 50% to perhaps 30–40% by a coincident decrease in CaCO₃ export that was modest enough not to violate observations of the glacial-age lysocline depth³³. Nevertheless, even a 30% increase in oceanic nutrients and in low- and mid-latitude export production would be a dramatic change.

Table 2 Atmospheric CO₂ effects of a 30% increase in the ocean nutrient inventory

	CaCO ₃ production	
	Proportional	Constant
Carbon cycle effect		
Closed system effects	–34	–43
Open system effects		
CaCO ₃ compensation	6	–3
Steady state lysocline shift	11	0
Total CO ₂ change (p.p.m.v.)	–17	–46

Given a residence time of 16 kyr or more for oceanic phosphate and few appropriate mechanisms for phosphate addition and removal³⁹, it is difficult to imagine how oceanic phosphate could vary so as to cause the observed amplitude, rate and phasing of atmospheric CO₂ change⁴⁰. On the other hand, there is growing evidence that the nitrogen cycle is adequately dynamic to allow for a large change in the oceanic nitrate reservoir on glacial/interglacial timescales⁴¹. Nitrogen isotope studies in currently active regions of denitrification^{42–44} indicate that water column denitrification was reduced in these regions during glacial periods. In addition, Falkowski⁴⁵ has suggested that N₂ fixation rates were greater during glacial periods because of increased airborne supply of iron-bearing dust to the open ocean (iron being an important requirement for N₂ fixation). Both of these hypothesized changes, a decrease in water column denitrification and an increase in N₂ fixation, would have increased the oceanic nitrate reservoir. It has been suggested that such a nitrate reservoir increase would lead to significantly higher export production in the open ocean during glacial periods, potentially explaining glacial/interglacial atmospheric CO₂ changes^{37,40,42,45}.

The major questions associated with this hypothesis involve the oceanic reservoir of phosphate, the feedbacks of the nitrogen cycle, and the strictness with which marine organisms must adhere to the Redfield ratios, both N/P and C/N. Oceanic N₂ fixation provides a mechanism to add fixed nitrogen to the ocean when the other nutrients are present in adequate supply. By contrast, riverine input of phosphorus, the major source of phosphate to the ocean, is not directly controlled by marine phytoplankton. As a result, geochemists have traditionally considered phosphate to be the fundamentally limiting major nutrient on glacial/interglacial timescales⁴⁶. For a nitrate reservoir increase to be the sole driver of an increase in low-latitude biological production, the nutrient requirements of the upper-ocean biota must be able to deviate from Redfield stoichiometry to adjust to changes in the N/P ratio of nutrient supply, so as to use the added nitrate fully in the absence of added phosphate. On the other hand, if this compensatory shift in biomass composition is incomplete, the surface ocean will tend to shift toward phosphate limitation.

It was originally suggested by Redfield³⁶, based on analogy with freshwater systems, that N₂ fixers in the open ocean will enjoy greater competitive success under conditions of nitrate limitation but will be discouraged in the case of phosphate limitation. If this sensitivity exists, then one would expect N₂ fixation to act as a negative feedback on the nitrate reservoir, varying so as to prevent large changes in the nitrate reservoir that are not associated with a coincident change in the phosphate reservoir⁴⁷. On the grand scale, this feedback must occur to prevent drastic swings in the oceanic nitrate reservoir. However, the quantitative constraint that this negative feedback places on the global nitrate reservoir is not yet known, with one suggested possibility being that iron is so important to N₂ fixers that changes in its supply can, within some limits, overpower the nitrate/phosphate sensitivity described by Redfield, leading to significant variations in the nitrate reservoir over time⁴⁵. Biogeochemical studies of the modern ocean⁴⁸ and the palaeoceanographic record⁴⁹ are both likely to play a role in the evaluation of these questions regarding Redfield stoichiometry and N₂ fixation, which are, in turn, critical for the hypothesis of the low-latitude biological pump as the driver of glacial/interglacial CO₂ change.

A number of palaeoceanographic studies suggest that export production was greater in low-latitude regions during ice ages^{50,51}. However, proxy data for nutrient utilization⁵² and pH⁵³ suggest that these changes were driven at least partially by enhanced upwelling, masking any potential sign of a change in the oceanic nutrient reservoir. Nutrient consumption is eventually complete in the modern low-latitude surface ocean, so that inorganic carbon is already stripped to the greatest possible degree, given the oceanic nutrient reservoir and the Redfield ratios. Vertical mixing and

upwelling supply excess inorganic carbon to the surface ocean in approximately Redfield proportions with the major nutrients, so that an increase in upwelling would increase the rate of export production from the surface ocean but would not enhance the extraction of DIC from the surface ocean. Increased export production driven by higher rates of vertical mixing or upwelling would only play a role in lowering atmospheric CO_2 if it caused a change in the chemical composition of the exported organic material (such as a decrease in its $\text{CaCO}_3/\text{C}_{\text{org}}$ ratio, as described above).

Nutrient utilization and high latitudes

Soon after Broecker's¹⁹ low-latitude-based hypothesis, several groups^{54–56} recognized the potential importance of the high-latitude surface ocean for glacial/interglacial CO_2 change, through its effect on the global efficiency of the biological pump. The thermocline of the world ocean outcrops at subpolar latitudes, allowing deep waters more ready access to the surface in the polar regions. In these regions, the nutrient-rich and CO_2 -charged waters of the deep ocean are exposed to the atmosphere and returned to the subsurface before the available nutrients are fully utilized by phytoplankton for carbon fixation. This incomplete use of upwelled nutrients allows for the leakage of deeply sequestered CO_2 back into the atmosphere, raising the atmospheric p_{CO_2} (Fig. 5). On this basis, increased nutrient utilization in the high-latitude surface ocean has been invoked as the cause of the lower atmospheric p_{CO_2} of glacial times^{16,54–58}.

With regard to their nutrient status, it can be shown that surface ocean regions that ventilate the ocean subsurface have particular importance for atmospheric CO_2 (Fig. 5). Briefly, this involves the integrated potential of global ocean nutrient stocks to drive carbon sequestration as the water holding these nutrients cycles through the surface ocean and drives biological export production. The 'Southern Ocean' (the high-latitude ocean of the Southern Hemisphere) ventilates large volumes of the ocean interior, making its variations in nutrient utilization a potentially important driver of glacial/interglacial CO_2 change (Fig. 6a). The Antarctic zone, the most polar region in the Southern Ocean, is central in the formation of deep and intermediate-depth waters (Fig. 6b), making this region important to the atmosphere/ocean CO_2 balance. The subantarctic zone (to the north of the Antarctic) has a less clear effect on atmospheric CO_2 , depending on the degree to which the nutrient status of the subantarctic surface influences the preformed nutrient concentration of newly formed subsurface water (Antarctic intermediate water and subantarctic mode water, Fig. 6b). In contrast to polar surface waters, the importance of the nutrient status of the low-latitude upwelling systems to atmospheric CO_2 does not extend to glacial/interglacial timescales because these regions do not ventilate the ocean interior. Put another way, nutrients upwelled in these environments are eventually consumed before the water holding them is folded back into the ocean interior, so nutrient utilization is complete today and would have been so during the last ice age as well.

In hypotheses of a change in Southern Ocean nutrient utilization, two causes have been considered: an increase in biological export production and a decrease in the exposure rate of deep waters at the polar surface. We might consider the first mechanism 'biological' and the second 'physical'. However, an increase in nutrient utilization by either mechanism requires certain sensitivities of the biological productivity in Southern Ocean surface waters. Specifically, both mechanisms would require that export production in the modern Southern Ocean is limited by something other than the major nutrients nitrate and phosphate that are supplied by the upwelling of deep water. Indeed, both light and trace metals such as iron are scarce commodities in the Southern Ocean³⁸. We can easily imagine processes that would have increased Southern Ocean nutrient utilization by either of the two mechanisms mentioned above. For instance, an increase in the input of dust and its

associated trace metals to the Southern Ocean might have driven an increase in the rate of nutrient and carbon uptake by phytoplankton³⁸. Alternatively, a strengthening of salinity-driven stratification in the Antarctic or a decrease in wind-driven upwelling in this region could have lowered the rate of nutrient and CO_2 transport to the surface⁵⁹.

With respect to rates of biological export production, palaeoceanographic work in the Southern Ocean indicates that the Antarctic and subantarctic have responded very differently to glacial/interglacial cycles. In the modern Antarctic, export production is dominated by diatoms, a group of phytoplankton that precipitate tests of hydrated silica, and these microfossils are frequently the dominant component of the underlying sediments. The accumulation of diatomaceous

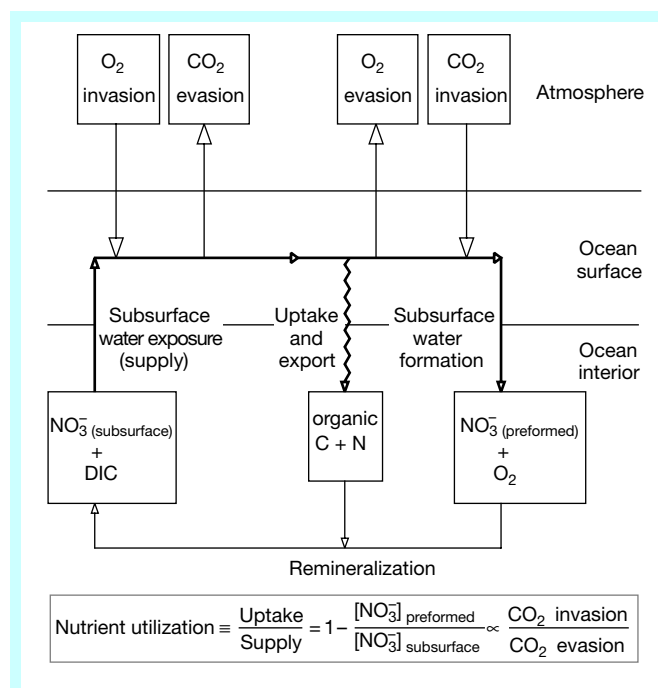


Figure 5 The effect on atmospheric CO_2 of the biological pump in a region of deep-ocean ventilation. Deep water is nutrient- and CO_2 -rich because of the lower-latitude biological pump, which sequesters both nutrients and inorganic carbon in the ocean interior; the major nutrients nitrate and phosphate are represented here as nitrate (NO_3^-) only. In high-latitude regions such as the Antarctic, the exposure of deep water at the surface releases this sequestered CO_2 to the atmosphere. However, the uptake of nutrients and CO_2 in the formation of phytoplankton biomass and the eventual export of organic matter subsequently lowers the p_{CO_2} of surface waters, causing the surface layer to reabsorb a portion of the CO_2 that was originally lost from the upwelled water. Thus, the ratio between export production and the ventilation of CO_2 -rich subsurface water, not the absolute rate of either process alone, controls the exchange of CO_2 between the atmosphere and high-latitude surface ocean. The nutrient concentration of water at the time that it enters the subsurface is referred to as its 'preformed' nutrient concentration. The preformed nutrient concentration of subsurface water provides an indicator of the cumulative nutrient utilization that this water underwent while at the surface, with a higher preformed nutrient concentration indicating lower cumulative nutrient utilization at the surface and thus a greater 'leak' in the biological carbon pump. The global efficiency of the biological pump can be evaluated in terms of the global mean preformed nutrient concentration of the ocean interior. In the calculation of this mean preformed nutrient concentration, each subsurface water formation term is weighted according to the volume of the ocean that it ventilates, because this corresponds to the amount of CO_2 that could be sequestered in a given portion of the ocean interior. As a result, surface ocean regions that are important in the ventilation of the ocean interior, such as the Antarctic, have particular importance for atmospheric CO_2 . This simple view overlooks some important facts. For instance, ocean/atmosphere CO_2 exchange is not instantaneous⁶⁸, and variations in the chemical composition of export production also affects the p_{CO_2} of surface waters, as does sea surface temperature.

sediments in the Antarctic was clearly lower during the last ice age⁶⁰, and palaeoceanographic indicators also suggest that the biological export of carbon was lower in general^{59,61,62}. However, given a possible bias in the proxies to record preferentially the carbon export associated with diatoms, it remains possible that the biological export of carbon from the glacial Antarctic surface was as high as or higher than export during interglacial times but that it occurred in a form that was poorly preserved in the sediment record^{64,65}. In the subantarctic, there is very strong evidence for higher export production during glacial times and an associated increase in the relative importance of diatom production^{59–61}. While these changes in export production are critical to the workings of the ice age Southern Ocean, they do not directly address the effect on atmospheric CO₂, which is largely determined by the history of nutrient utilization (Fig. 5).

In the Antarctic, nitrogen isotope data suggest that nitrate utilization during the last ice age was twice or more its current value^{59,66}. These results seem to support the longstanding hypothesis that nutrient utilization changes in Antarctic waters are a fundamental driver of glacial/interglacial changes in atmospheric pCO₂. A box model calculation predicts that 25–40% higher nitrate utilization (that is, 50–65% during the last ice age compared to 25% during the present interglacial) could lower atmospheric CO₂ by the full glacial/interglacial amplitude⁶⁶. According to the model calculation, a significant part of this CO₂ change is due to an increase in ocean alkalinity, and this model calculation may underestimate its importance⁶⁷. In addition, a glacial increase in wintertime sea ice coverage may have worked to lower atmospheric CO₂ by preventing gas exchange during this season⁶⁸, when Antarctic surface water is more likely to be supersaturated with CO₂. Because palaeoceanographic proxy data suggest that Antarctic export production was lower during the last ice age, we infer that more complete nitrate utilization in the Antarctic was due to a lower rate of nitrate supply from the subsurface, implying that the fundamental driver of the CO₂ change was an ice age decrease in the ventilation of deep waters at the surface of the Antarctic⁵⁹.

The interpretation from nitrogen isotopes of higher nitrate utilization in the Antarctic faces a number of apparent disagreements with other proxies of nutrient status, in particular, the Cd/Ca (refs 69, 70) and ¹³C/¹²C (ref. 71) of planktonic foraminiferal calcite. The Cd/Ca of surface water varies predictably with its nutrient content^{69,72}. The same is roughly true of the ¹³C/¹²C of surface DIC, although with the additional influence of gas exchange⁷³. Measurements of these ratios in planktonic foraminifera indicate no clear decrease in the nutrient concentration of Antarctic surface water during the last ice age, while such a decrease would have been expected on the basis of the nitrogen isotope data, assuming no major changes in the nutrient concentration of the Antarctic subsurface (Fig. 5). Unfortunately, each of the palaeochemical proxies has significant uncertainties. In the case of nitrogen isotopes, concerns currently revolve around the controls on the isotope fractionation of nitrate uptake by phytoplankton^{74,75}. Foraminiferal proxies for surface conditions also have potential complications^{13,72,76}, with the water depth of shell growth⁷⁷ being one of the greatest concerns for the Antarctic, a region with the potential for very sharp vertical gradients within the upper water column.

Dissolved silicate is a major nutrient for diatom growth because of the silica tests that these phytoplankton precipitate. Much like nitrate and phosphate, silicate is nearly completely depleted in the low-latitude surface ocean but is found at higher concentrations in the Antarctic. The silicon isotopic composition of diatom microfossils implies that there was a reduction in Antarctic silicate utilization during the last ice age⁷⁸, in contrast to the evidence for enhanced nitrate utilization. While this may indicate a disagreement among proxies, the alternative is that the difference signals a real oceanographic change. Field observations, incubations and culture studies^{79,80} indicate that iron-replete conditions (such as may have

resulted from the dustiness of the ice age atmosphere⁸¹) favour a higher nitrate/silicate uptake ratio in diatoms. Taking both the nitrogen and silicon isotope data at face value implies that the nitrate/silicate uptake ratio of phytoplankton was higher during the last ice age, leading to lower nitrate but higher silicate concentrations in the glacial Antarctic surface relative to modern times⁷⁸. If the change in uptake ratio was great enough, it may actually have removed the tendency for preferential depletion of silicate relative to nitrate that is observed in the modern Southern Ocean surface.

In the subantarctic, planktonic foraminiferal Cd/Ca and the ¹³C/¹²C of diatom-bound organic matter both show changes that are consistent with an ice age state of higher nutrient utilization^{62,82}. However, the consideration of proxy complications may argue against such a change⁷². The significance of the nitrogen isotope data for the glacial subantarctic is uncertain. On the one hand, bulk sediment nitrogen isotope records from the subantarctic can be interpreted as suggesting lower utilization during the last ice age, perhaps due to an increase in the supply of nutrients from below⁵⁹. On the other hand, isotopic measurements of nitrate in the modern Southern Ocean indicate that downcore nitrogen isotope changes in the subantarctic may instead represent a shift in the nitrate source to the subantarctic surface, with the Antarctic surface providing most of this nitrate during interglacial times, but the subantarctic thermocline becoming the dominant nitrate source during glacial periods⁷⁵. This latter interpretation has some merit in its consistency with the Antarctic nitrogen isotope data. If Antarctic nitrate utilization was higher during glacial times as suggested above, then the Antarctic surface would have had less nitrate to transport into the subantarctic surface, making other sources of nitrate proportionally more important.

In spite of the many uncertainties and conflicts, we propose a hypothesis for the Southern Ocean during the last ice age which would explain lower glacial atmospheric CO₂ concentrations through reduced deep-water ventilation and enhanced nutrient utilization (Fig. 6c and d). In this hypothesis, a cooler climate caused a northward shift in the belt of eastward winds that drives upwelling and northward surface flow in the modern Antarctic. This shift caused a decrease in the upwelling of deep water into the Antarctic surface⁸³, replacing it with upwelling of intermediate-depth water into the subantarctic surface. The hypothetical glacial Antarctic, because of reduced wind-driven upwelling, was able to develop a stable, fresh, frequently ice-covered surface layer⁸⁴, further reducing deep-ocean ventilation in the open Antarctic. These changes in circulation resulted in a lower rate of nutrient supply to the Antarctic surface⁵⁹ and reduced CO₂ outgassing^{67,68}.

From interglacial to glacial times, export production in the Antarctic decreased less than did the supply of nitrate and phosphate from below, that is, nutrient utilization increased during glacial times⁶⁶. This response may indicate that the iron limitation of modern Antarctic phytoplankton was alleviated during glacial times by the increase in dust deposition from the continents^{58,81}. Dissolved silicate may have been decoupled from nitrate and phosphate changes⁷⁸, owing to more frugal use of silicate by diatoms^{79,80} and the possible importance of non-diatom phytoplankton in the glacial Antarctic⁶⁴.

The hypothetical glacial subantarctic was more productive, perhaps also because of an increased supply of iron from dust⁶¹. Subantarctic utilization of nitrate and phosphate may have increased^{62,82} due to a combination of higher export production and a reduction in the supply of these nutrients from the Antarctic surface. However, increased nutrient supply from the thermocline below the subantarctic surface may have worked against such a change⁵⁹.

Higher utilization of nitrate and phosphate across the Antarctic and perhaps also the subantarctic lowered atmospheric pCO₂, with the Antarctic dominating this effect because of the large volume of the ocean that it currently ventilates⁶⁷. The reduced supply of nitrate

and phosphate from the Antarctic to the subantarctic represented a decrease in one of the most important routes by which these nutrients are fed from the cold, deep ocean into the mid-depth and upper ocean. In addition, the high export production of the glacial subantarctic may have worked to pump nutrients and DIC out of upper and mid-depth ocean by producing rapidly sinking particles that persist into the deep ocean (Fig. 6d). As a result, the nutrient content of the mid-depth ocean decreased while its oxygen content increased^{16,42–44,85–87}. This change was balanced by an increase in nutrient content and a decrease in oxygen content in the abyssal ocean^{59,85}, as well as a calcite dissolution event in the abyssal ocean at the onset of ice ages, which further reduced the

atmospheric p_{CO_2} (refs 66, 67, 88). Since North Atlantic Deep Water formation and Antarctic upwelling are coupled in the modern ocean^{89,90}, a glacial reduction in Antarctic upwelling is consistent with the evidence for less North Atlantic Deep Water formation during the last ice age⁹¹. The proposed upwelling of intermediate-depth waters into the glacial subantarctic surface is akin to a shoaling of modern Southern Ocean overturning, with new intermediate-depth waters being supplied from the North Atlantic⁹⁹ and/or North Pacific¹⁰⁰ (Fig. 6c).

This Southern Ocean hypothesis is controversial (for example, see ref. 72). Clearly, one of the major obstacles in the evaluation of this and other hypotheses is the uncertainty in our proxies for surface-

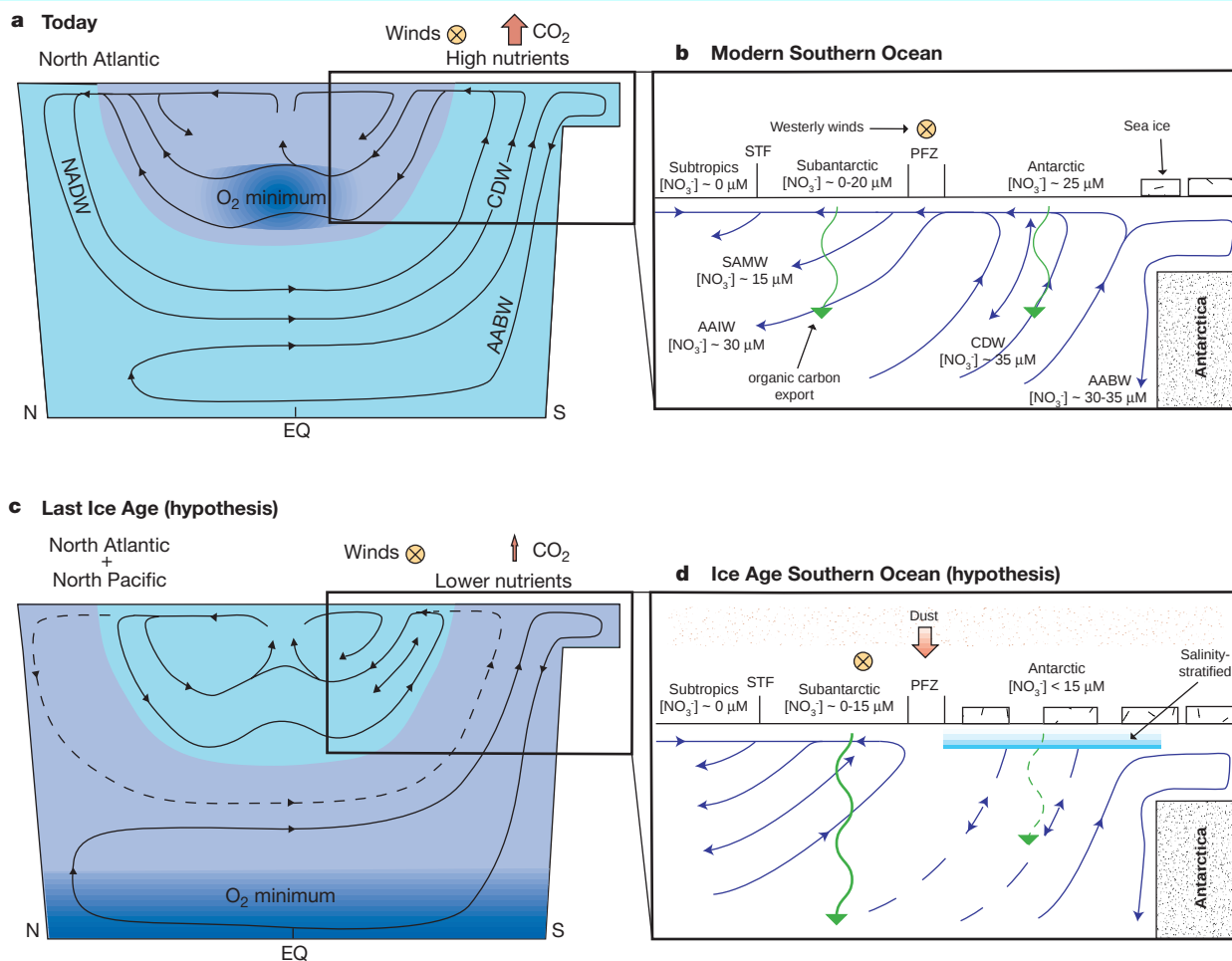


Figure 6 The modern ocean (**a, b**) and a Southern Ocean-based hypothesis for reduced levels of atmospheric CO₂ during glacial times (**c, d**). The figure shows a generalized depth section running from north to south (**a, c**) with an expanded view of the Southern Ocean (**b, d**). The north–south and vertical components of ocean circulation are shown (black flow lines in **a** and **c**, blue in **b** and **d**). The blue shading in **a** and **c** indicates changes in the intensity of circulation between modern and ice-age times, with the lighter shade representing more vigorous circulation, illustrating the hypothesis that ice-age circulation was less vigorous in the cold, dense, deep ocean but more vigorous in the warmer, less dense, upper ocean. In the modern ocean (**a, b**), circulation of the interior can be characterized as follows. New deep water forms in the high-latitude North Atlantic (North Atlantic Deep Water, NADW). NADW flows south and mixes with other deep waters (including Antarctic Bottom Water, AABW) to generate Circumpolar Deep Water (CDW). Eastward winds (orange circled cross) over the Polar Frontal Zone (PFZ) drive upwelling of CDW into the Antarctic surface, releasing deep-sequestered CO₂ to the atmosphere (upward red arrow in **a**). Antarctic surface water flows northward to the PFZ, where it sinks into the subsurface as Antarctic Intermediate Water (AAIW) or mixes with subantarctic surface water, some of which forms Subantarctic Mode Water (SAMW). The biological production and export of organic matter (green arrows) extracts nutrients and inorganic carbon from surface waters, causing the nitrate concentration ([NO₃]) of the Antarctic surface (~25 μM), the subantarctic surface (0–20 μM), and newly formed AAIW and SAMW (~30 and ~15 μM, respectively) to be lower than that of the CDW that was originally upwelled (~35 μM). AAIW and SAMW ventilate the mid-depth ocean and supply nutrients to the lower-latitude surface. This nutrient supply fuels biological export production, the degradation of which causes a subsurface deficit in dissolved oxygen which is most intense at intermediate depths (the O₂ minimum in **a**). The hypothetical ice-age Southern Ocean described in the text includes the following changes (**c, d**). Upwelling into the Antarctic surface decreased, possibly due to an equatorward shift in the belt of eastward winds, and/or a freshening of the surface ('salinity-stratification') related to an increase in sea ice. In the subantarctic, the winds may have driven the upwelling of intermediate-depth waters, which formed in the North Atlantic and/or North Pacific. The supply of iron from dust was high (broad downward arrows and red stippling). Organic carbon export was higher in the subantarctic but lower in the Antarctic (green arrows). Possibly due to the greater supply of dust-borne iron, the degree of nitrate utilization was higher in both regions, leading to lower surface [NO₃] and less nutrient transport to the lower latitudes. This pattern of nutrient supply and export production caused the O₂ minimum to migrate into the abyssal ocean and reduced the release of CO₂ from the Southern Ocean.

ocean nutrient status and export production. However, more global properties such as the $^{13}\text{C}/^{12}\text{C}$ ratio of atmospheric CO_2 (refs 92, 93), the O_2 content of the ocean subsurface, and the relative timing of geochemical and climatic changes⁴⁰ are also important constraints on this and other hypotheses for glacial/interglacial CO_2 change. For instance, atmospheric CO_2 is among the early parameters to change going into an interglacial period, nearly coincident with the warming of high southern latitudes⁹⁴, which supports a Southern Hemisphere origin for glacial/interglacial CO_2 change.

Conclusion

The most viable hypotheses for the cause of glacial/interglacial CO_2 change involve the extraction of carbon from the surface ocean by biological production, either at low or high latitudes, necessarily allied with changes in the marine calcium carbonate budget. Much work remains to be done before a consensus can emerge on the specific driver of glacial/interglacial CO_2 changes. While our ability to reconstruct past environmental conditions requires continued effort, there also remain fundamental uncertainties about the sensitivity of atmospheric CO_2 to the oceanic changes discussed above¹⁷. Nevertheless, our understanding of the potential causes is much greater than when the CO_2 changes were first discovered.

Lessons learned from the investigation of glacial/interglacial CO_2 change inform our thinking on the critical issues of the modern carbon cycle. Several of the hypotheses for glacial/interglacial CO_2 change relate directly to considerations for purposeful mitigation of the anthropogenic CO_2 increase^{45,58}. Other hypothetical causes for glacial/interglacial CO_2 change have been recognized as potentially important factors in possible situations of anthropogenic CO_2 increase and global warming⁹⁵. More generally, the mystery of glacial/interglacial CO_2 change has provided a focus for study of the interaction among the diverse processes operating in the natural environment, resulting in real progress toward a science of the Earth system. □

- Hays, J. D., Imbrie, J. & Shackleton, N. J. Variations in the Earth's orbit: Pacemaker of the Ice Ages. *Science* **194**, 1121–1132 (1976).
- Berger, A., Imbrie, J., Hays, J., Kukla, G. & Saltzman, B. (eds) *Milankovitch and Climate* (Reidel, Boston, 1984).
- Barnola, J. M., Raynaud, D., Korotkevich, Y. S. & Lorius, C. Vostok ice core provides 160,000-year record of atmospheric CO_2 . *Nature* **329**, 408–414 (1987).
- Petit, J. R. *et al.* Climate and atmospheric history of the past 420,000 years from the Vostok ice core, Antarctica. *Nature* **399**, 429–436 (1999).
- Webb, R. S., Lehman, S. J., Rind, D. H., Healy, R. J. & Sigman, D. M. Influence of ocean heat transport on the climate of the Last Glacial Maximum. *Nature* **385**, 695–699 (1997).
- Broecker, W. S. Glacial to interglacial changes in ocean chemistry. *Progr. Oceanogr.* **2**, 151–197 (1982).
- Adams, J. M., Faure, H., Faure-Denard, L., McGlade, J. M. & Woodward, F. I. Increases in terrestrial carbon storage from the Last Glacial Maximum to the present. *Nature* **348**, 711–714 (1990).
- Crowley, T. J. Ice-Age terrestrial carbon changes revisited. *Glob. Biogeochem. Cycles* **9**, 377–389 (1995).
- Shackleton, N. J. in *The Fate of Fossil Fuel CO_2 in the Oceans* (eds Sundquist, E. T. & Broecker, W. S.) 401–427 (American Geophysical Union, Washington DC, 1977).
- Curry, W. B., Duplessy, J. C., Labeyrie, L. D. & Shackleton, N. J. Changes in the distribution of $\delta^{13}\text{C}$ of deep water TCO_2 between the last glaciation and the Holocene. *Paleoceanography* **3**, 317–341 (1988).
- Duplessy, J. C. *et al.* Deepwater source variations during the last climatic cycle and their impact on the global deepwater circulation. *Paleoceanography* **3**, 343–360 (1988).
- Bird, M. I., Lloyd, J. & Farquhar, G. D. Terrestrial carbon storage at the LGM. *Nature* **371**, 566 (1994).
- Spero, H., Bijma, J., Lea, D. & Bemis, B. Effect of seawater carbonate concentration on foraminiferal carbon and oxygen isotopes. *Nature* **390**, 497–500 (1997).
- CLIMAP. The surface of the ice-age earth. *Science* **191**, 1131–1144 (1976).
- Guilderson, T. P., Fairbanks, R. G. & Rubenstone, J. L. Tropical temperature variations since 20,000 years ago: modulating interhemispheric climate change. *Science* **263**, 663–665 (1994).
- Keir, R. S. On the late Pleistocene ocean geochemistry and circulation. *Paleoceanography* **3**, 413–445 (1988).
- Broecker, W. S. *et al.* How strong is the Harvardton-Bear constraint? *Glob. Biogeochem. Cycles* **13**, 817–820 (1999).
- Fairbanks, R. G. A 17,000-year glacio-eustatic sea level record: Influence of glacial melting rates on the Younger Dryas event and deep-ocean circulation. *Nature* **342**, 637–642 (1989).
- Broecker, W. S. Ocean chemistry during glacial time. *Geochim. Cosmochim. Acta* **46**, 1689–1706 (1982).
- Berger, W. H. Increase of carbon dioxide in the atmosphere during deglaciation: The coral reef hypothesis. *Naturwissenschaften* **69**, 87–88 (1982).
- Broecker, W. S. & Peng, T.-H. The role of CaCO_3 compensation in the glacial to interglacial

- atmospheric CO_2 change. *Glob. Biogeochem. Cycles* **1**, 15–29 (1987).
- Milliman, J. D. *Marine Carbonates* (Springer, New York, 1974).
- Edmond, J. M. *et al.* Ridge crest hydrothermal activity and the balances of the major and minor elements in the ocean: The Galapagos data. *Earth Planet. Sci. Lett.* **46**, 1–18 (1979).
- Sayles, F. L. & Mangelsdorf, P. C. Jr The equilibration of clay minerals with seawater: exchange reactions. *Geochim. Cosmochim. Acta* **41**, 951–960 (1977).
- Berger, W. H. Planktonic foraminifera: selective solution and the lysocline. *Mar. Geol.* **8**, 111–138 (1970).
- Archer, D. Modeling the calcite lysocline. *J. Geophys. Res.* **96**, 17037–17050 (1991).
- Broecker, W. S. & Takahashi, T. The relationship between lysocline depth and *in situ* carbonate ion concentration. *Deep-Sea Res.* **25**, 65–95 (1978).
- Broecker, W. S. in *The Late Cenozoic Glacial Ages* (ed. Turekian, K. K.) 239–265 (Yale Univ. Press, New Haven, Connecticut, 1971).
- Catubig, N. *et al.* Global deep-sea burial rate of calcium carbonate during the last glacial maximum. *Paleoceanography* **13**, 298–310 (1998).
- Archer, D. & Maier-Raimer, E. Effect of deep-sea sedimentary calcite preservation on atmospheric CO_2 concentration. *Nature* **367**, 260–263 (1994).
- Dymond, J. & Lyle, M. Flux comparisons between sediments and sediment traps in the eastern tropical Pacific: Implications for atmospheric CO_2 variations during the Pleistocene. *Limnol. Oceanogr.* **30**, 699–712 (1985).
- Keir, R. S. Is there a component of Pleistocene CO_2 change associated with carbonate dissolution cycles? *Paleoceanography* **10**, 871–880 (1995).
- Sigman, D. M., McCorkle, D. C. & Martin, W. R. The calcite lysocline as a constraint on glacial/interglacial low-latitude production changes. *Glob. Biogeochem. Cycles* **12**, 409–427 (1998).
- Emerson, S. & Bender, M. L. Carbon fluxes at the sediment-water interface of the deep sea: calcium carbonate preservation. *J. Mar. Res.* **39**, 139–162 (1981).
- Hales, B. & Emerson, S. Calcite dissolution in sediments of the Ceara Rise: In situ measurements of porewater O_2 , pH, and $\text{CO}_2(\text{aq})$. *Geochim. Cosmochim. Acta* **61**, 501–514 (1997).
- Redfield, A. C., Ketchum, B. H. & Richards, F. A. in *The Sea* (ed. Hill, M. N.) Vol. 2, 26–77 (Interscience, New York, 1963).
- McElroy, M. B. Marine biological controls on atmospheric CO_2 and climate. *Nature* **302**, 328–329 (1983).
- Chisholm, S. W. & Morel, F. M. M. (eds) What controls phytoplankton production in nutrient-rich areas of the open sea? *Limnol. Oceanogr.* **36**(8) (special volume) 1507–1970 (1991).
- Ruttenberg, K. C. Reassessment of the oceanic residence time of phosphorous. *Chem. Geol.* **107**, 405–409 (1993).
- Broecker, W. S. & Henderson, G. M. The sequence of events surrounding Termination II and their implications for the cause of glacial-interglacial CO_2 changes. *Paleoceanography* **13**, 352–364 (1998).
- Gruher, N. & Sarmiento, J. L. Global patterns of marine nitrogen fixation and denitrification. *Glob. Biogeochem. Cycles* **11**, 235–266 (1997).
- Ganeshran, R. S., Pedersen, T. F., Calvert, S. E. & Murray, J. W. Large changes in oceanic nutrient inventories from glacial to interglacial periods. *Nature* **376**, 755–758 (1995).
- Altabet, M. A., Francois, R., Murray, D. W. & Prell, W. L. Climate-related variations in denitrification in the Arabian Sea from sediment $^{15}\text{N}/^{14}\text{N}$ ratios. *Nature* **373**, 506–509 (1995).
- Pride, C. *et al.* Nitrogen isotopic variations in the Gulf of California since the last deglaciation: Response to global climate change. *Paleoceanography* **14**, 397–409 (1999).
- Falkowski, P. G. Evolution of the nitrogen cycle and its influence on the biological sequestration of CO_2 in the ocean. *Nature* **387**, 272–275 (1997).
- Broecker, W. S. & Peng, T.-H. *Tracers in the Sea* (Eldigio, Palisades, New York, 1982).
- Tyrrel, T. The relative influences of nitrogen and phosphorus on oceanic primary production. *Nature* **400**, 525–531 (1999).
- Karl, D. *et al.* The role of nitrogen fixation in biogeochemical cycling in the subtropical North Pacific Ocean. *Nature* **388**, 533–538 (1997).
- Haug, G. H. *et al.* Glacial/interglacial variations in productivity and nitrogen fixation in the Cariaco Basin during the last 550 ka. *Paleoceanography* **13**, 427–432 (1998).
- Pedersen, T. F. Increased productivity in the eastern equatorial Pacific during the last glacial maximum (19,000 to 14,000 yr B. P.). *Geology* **11**, 16–19 (1983).
- Berger, W. H., Herguera, J. C., Lange, C. B. & Schneider, R. in *Carbon Cycling in the Glacial Ocean: Constraints on the Ocean's Role in Global Change* (eds Zahn, R., Kaminski, M., Labeyrie, L. & Pederson, T. F.) 385–412 (Springer, New York, 1994).
- Farrell, J. W., Pedersen, T. F., Calvert, S. E. & Nielsen, B. Glacial–interglacial changes in nutrient utilization in the equatorial Pacific Ocean. *Nature* **377**, 514–517 (1995).
- Sanyal, A., Hemming, N. G., Broecker, W. S. & Hanson, G. N. Changes in pH in the eastern equatorial Pacific across stage 5–6 boundary based on boron isotopes in foraminifera. *Glob. Biogeochem. Cycles* **11**, 125–133 (1997).
- Knox, F. & McElroy, M. Changes in atmospheric CO_2 influence of the marine biota at high latitude. *J. Geophys. Res.* **89**, 4629–4637 (1984).
- Sarmiento, J. L. & Toggweiler, J. R. A new model for the role of the oceans in determining atmospheric pCO_2 . *Nature* **308**, 621–624 (1984).
- Siegenthaler, U. & Wenk, T. Rapid atmospheric CO_2 variations and ocean circulation. *Nature* **308**, 624–626 (1984).
- Broecker, W. S. & Peng, T.-H. The cause of the glacial to interglacial atmospheric CO_2 change: A polar alkalinity hypothesis. *Glob. Biogeochem. Cycles* **3**, 215–239 (1989).
- Martin, J. H. Glacial-interglacial CO_2 change: The iron hypothesis. *Paleoceanography* **5**, 1–13 (1990).
- Francois, R. F. *et al.* Water column stratification in the Southern Ocean contributed to the lowering of glacial atmospheric CO_2 . *Nature* **389**, 929–935 (1997).
- Mortlock, R. A. *et al.* Evidence for lower productivity in the Antarctic during the last glaciation. *Nature* **351**, 220–223 (1991).
- Kumar, N. *et al.* Increased biological productivity and export production in the glacial Southern Ocean. *Nature* **378**, 675–680 (1995).
- Rosenthal, Y., Daham, M. & Shemesh, A. Southern Ocean contributions to glacial-interglacial changes of atmospheric CO_2 : Evidence from carbon isotope records in diatoms. *Paleoceanography* **15**, 65–75 (2000).
- Smith, H. J., Fischer, H., Wahlen, M., Mastroian, D. & Deck, B. Dual modes of the carbon cycle since the Last Glacial Maximum. *Nature* **400**, 248–250 (1999).

64. Di Tullio, G. R. *et al.* Rapid and early export of Phaeocystis antarctica blooms in the Ross Sea, Antarctica. *Nature* **404**, 595–598 (2000).
65. Moore, J. K., Abbott, M. R., Richman, J. G. & Nelson, D. M. The Southern Ocean at the last glacial maximum: A strong sink for atmospheric carbon dioxide. *Glob. Biogeochem. Cycles* **14**, 455–475 (2000).
66. Sigman, D. M., Altabet, M. A., Francois, R., McCorkle, D. C. & Gaillard, J.-F. The isotopic composition of diatom-bound nitrogen in Southern Ocean sediments. *Paleoceanography* **14**, 118–134 (1999).
67. Toggweiler, J. R. Variations in atmospheric CO₂ driven by ventilation of the ocean's deepest water. *Paleoceanography* **14**, 571–588 (1999).
68. Stephens, B. B. & Keeling, R. F. The influence of Antarctic sea ice on glacial/interglacial CO₂ variations. *Nature* **404**, 171–174 (2000).
69. Boyle, E. A. Cadmium: Chemical tracer of deepwater paleoceanography. *Paleoceanography* **3**, 471–489 (1988).
70. Keigwin, L. D. & Boyle, E. A. Late quaternary paleochemistry of high-latitude surface waters. *Paleogeogr. Palaeoclim. Paleocool.* **73**, 85–106 (1989).
71. Charles, C. D. & Fairbanks, R. G. in *Geological History of the Polar Oceans: Arctic versus Antarctic* (eds Bleil, U. & Thiede, J.) 519–538 (Kluwer Academic, Boston, 1988).
72. Elderfield, H. & Rickaby, R. E. M. Oceanic Cd/P ratio and nutrient utilization in the glacial Southern Ocean. *Nature* **405**, 305–310 (2000).
73. Broecker, W. S. & Maier-Reimer, E. The influence of air and sea exchange on the carbon isotope distribution in the sea. *Glob. Biogeochem. Cycles* **6**, 315–320 (1992).
74. Altabet, M. A. & Francois, R. Sedimentary nitrogen isotopic ratio as a recorder for surface ocean nitrate utilization. *Glob. Biogeochem. Cycles* **8**, 103–116 (1994).
75. Sigman, D. M., Altabet, M. A., Francois, R., McCorkle, D. C. & Fischer, G. The $\delta^{15}\text{N}$ of nitrate in the Southern Ocean: Consumption of nitrate in surface waters. *Glob. Biogeochem. Cycles* **13**, 1149–1166 (1999).
76. McCorkle, D. C., Martin, P. A., Lea, D. W. & Klinkhammer, G. P. Evidence of a dissolution effect on benthic foraminiferal shell chemistry: $\delta^{13}\text{C}$, Cd/Ca, Ba/Ca, and Sr/Ca results from the Ontong Java plateau. *Paleoceanography* **10**, 699–714 (1995).
77. Kohfeld, K. E., Fairbanks, R. G., Smith, S. L. & Walsh, I. D. *Neogloboquadrina pachyderma* (sinistral coiling) as paleoceanographic tracers in polar oceans: Evidence from northeast water polynya plankton tows, sediment traps, and surface sediments. *Paleoceanography* **11**, 679–699 (1996).
78. De La Rocha, C. L., Brzezinski, M. A., DeNiro, M. J. & Shemesh, A. Silicon-isotope composition of diatoms as an indicator of past oceanic change. *Nature* **395**, 680–683 (1998).
79. Hutchins, D. A. & Bruland, K. W. Iron-limited diatom growth and Si:N uptake ratios in a coastal upwelling regime. *Nature* **393**, 561–564 (1998).
80. Takeda, S. Influence of iron availability on nutrient consumption ratio of diatoms in oceanic waters. *Nature* **393**, 774–777 (1998).
81. Mahowald, N. *et al.* Dust sources and deposition during the last glacial maximum and current climate: A comparison of model results with paleodata from ice cores and marine sediments. *J. Geophys. Res. Atmos.* **104**, 15895–15916 (1998).
82. Rosenthal, Y., Boyle, E. A. & Labeyrie, L. Last glacial maximum paleochemistry and deepwater circulation in the Southern Ocean: evidence from foraminiferal cadmium. *Paleoceanography* **12**, 787–796 (1997).
83. Toggweiler, J. R., Carson, S. & Bjornsson, H. Response of the ACC and the Antarctic pycnocline to a meridional shift in the southern hemisphere westerlies. *Eos* **80**, OS286 (1999).
84. Morley, J. J. & Hays, J. D. Oceanographic conditions associated with high abundances of the radiolarian *Cycladophora davisiana*. *Earth Planet. Sci. Lett.* **66**, 63–72 (1983).
85. Sarmiento, J. L. & Orr, J. C. Three-dimensional simulations of the impact of Southern Ocean nutrient depletion on atmospheric CO₂ and ocean chemistry. *Limnol. Oceanogr.* **36**, 1928–1950 (1991).
86. Herguera, J. C., Jansen, E. & Berger, W. H. Evidence for a bathyal front at 2000 m depth in the glacial Pacific, based on a depth transect on Ontong Java Plateau. *Paleoceanography* **7**, 273–288 (1992).
87. Behl, R. J. & Kennett, J. P. Brief interstadial events in the Santa Barbara basin, NE Pacific, during the past 60 kyr. *Nature* **379**, 243–246 (1996).
88. Boyle, E. A. The role of vertical chemical fractionation in controlling late quaternary atmospheric carbon dioxide. *J. Geophys. Res.* **93**, 15701–15714 (1988).
89. Toggweiler, J. R. & Samuels, B. Effect of Drake Passage on the global thermohaline circulation. *Deep-Sea Res.* **142**, 477–500 (1998).
90. Gnanadesikan, A. A simple predictive model for the structure of the oceanic pycnocline. *Science* **283**, 2077–2079 (1999).
91. Boyle, E. A. & Keigwin, L. D. Deep circulation of the North Atlantic for the last 200,000 years: geochemical evidence. *Science* **218**, 784–787 (1982).
92. Leuenberger, M., Siegenthaler, U. & Langway, C. C. Carbon isotope composition of atmospheric CO₂ during the last Ice Age from an Antarctic ice core. *Nature* **357**, 488–490 (1992).
93. Marino, B. D. & McElroy, M. B. Isotopic composition of atmospheric CO₂ inferred from carbon in C4 plant cellulose. *Nature* **349**, 127–131 (1991).
94. Sowers, T. & Bender, M. Climate records covering the last deglaciation. *Science* **269**, 210–214 (1995).
95. Sarmiento, J. L., Hughes, T. M. C., Stouffer, R. J. & Manabe, S. Response of the ocean carbon cycle to anthropogenic climate warming. *Nature* **393**, 245–249 (1998).
96. Bassinot, F. C. *et al.* The astronomical theory of climate and the age of the Brunhes-Matuyama magnetic reversal. *Earth Planet. Sci. Lett.* **126**, 91–108 (1994).
97. Holmen, K. in *Global Biogeochemical Cycles* (eds Butcher, S. S., Charlson, R. J., Orians, G. H. & Wolfe, G. V.) 239–262 (Academic, New York, 1992).
98. Honjo, S. in *Particle Flux in the Ocean* (eds Ittekkot, V., Schafer, P., Honjo, S. & Depetris, P. J.) (Wiley Interscience, Munich, 1996).
99. Oppo, D. W. & Lehman, S. J. Mid-depth circulation of the subpolar North Atlantic during the last glacial maximum. *Science* **259**, 1148–1152 (1993).
100. Keigwin, L. D. Glacial-age hydrography of the far northwest Pacific Ocean. *Paleoceanography* **13**, 323–339 (1998).

Acknowledgements

We thank R. F. Anderson, M. L. Bender, M. A. Brzezinski and J. R. Toggweiler for discussions. We are indebted to P. G. Falkowski, G.M. Henderson and C. Prentice for comments on the manuscript.

Correspondence and requests for materials should be addressed to D.M.S. (e-mail: sigman@princeton.edu).



Vasoregulation by the $\beta 1$ subunit of the calcium-activated potassium channel

Robert Brenner*, Guillermo J. Pérez†, Adrian D. Bonev†, Delrae M. Eckman†, Jon C. Kosek‡, Steven W. Wiler*, Andrew J. Patterson§, Mark T. Nelson† & Richard W. Aldrich*

* Department of Molecular and Cellular Physiology and Howard Hughes Medical Institute, Stanford University, Stanford, California 94305, USA

† Department of Pharmacology, College of Medicine, The University of Vermont, Burlington, Vermont 05405, USA

‡ Department of Pathology, Palo Alto Veterans Administration Healthcare System, Palo Alto, California 94305 and Stanford University School of Medicine, Stanford, California 94305, USA

§ Department of Anesthesia, Stanford University School of Medicine, Stanford, California 94305, USA

Small arteries exhibit tone, a partially contracted state that is an important determinant of blood pressure. In arterial smooth muscle cells, intracellular calcium paradoxically controls both contraction and relaxation. The mechanisms by which calcium can differentially regulate diverse physiological responses within a single cell remain unresolved. Calcium-dependent relaxation is mediated by local calcium release from the sarcoplasmic reticulum. These 'calcium sparks' activate calcium-dependent potassium (BK) channels comprised of α and $\beta 1$ subunits. Here we show that targeted deletion of the gene for the $\beta 1$ subunit leads to a decrease in the calcium sensitivity of BK channels, a reduction in functional coupling of calcium sparks to BK channel activation, and increases in arterial tone and blood pressure. The $\beta 1$ subunit of the BK channel, by tuning the channel's calcium sensitivity, is a key molecular component in translating calcium signals to the central physiological function of vasoregulation.

Chronic blood pressure elevation resulting from increased arterial tone creates a burden on many organs leading to stroke, heart disease and renal disease¹. An understanding of the molecules involved in regulating arterial tone is crucial to improving our understanding of essential hypertension and perhaps to developing better therapies. Large-conductance calcium-activated potassium channels (BK channels) are pivotal in the regulation of arterial tone, where they facilitate a negative feedback mechanism which opposes vasoconstriction². Intravascular pressure increases arterial tone by a complex process which includes a graded membrane depolarization and elevation of calcium influx through dihydropyridine-sensitive, voltage-dependent calcium channels^{3–5}. Calcium influx causes a global increase in cytoplasmic calcium leading to vasoconstriction. Calcium influx also activates localized calcium release events from ryanodine receptors, termed calcium sparks, which in turn activate nearby calcium-activated potassium (BK) channels, causing a hyperpolarizing current to oppose vasoconstriction⁶. Arterial tone therefore results from the interplay of opposing calcium-dependent processes: constriction, which is driven by global increases in calcium; and relaxation, which is driven by localized calcium concentrations². The important role of BK channels in smooth muscle is demonstrated when the channels are specifically blocked with iberiotoxin³, which leads to marked membrane depolarization and vasoconstriction^{4,5} and inhibits the actions of a variety of smooth muscle relaxants^{2,6}.

BK channels are broadly expressed, and have functional roles in vascular smooth muscle as well as other tissues including skeletal muscle, neurons, kidney and secretory cells^{7–10}. The functional diversity required for the tissue-specific roles of BK channels may be created in part by association with accessory β -subunits. A family of four BK β -subunits has been identified^{11–20}. Each family member has a different tissue distribution and different effects on BK channel pharmacology and activation gating. The $\beta 1$ -subunit is enriched in smooth muscle and purifies with the BK pore-forming subunit²¹. In expression systems, the $\beta 1$ subunit confers an increased calcium sensitivity, slows gating kinetics and increases

the sensitivity to the agonist dehydrosoyasaponin (DHS-1)^{20,22–24}. Using sensitivity to DHS-1 as a probe for BK $\alpha/\beta 1$ subunits, it has been shown that human coronary artery smooth muscle is enriched for $\alpha/\beta 1$ -assembled BK channels, and that these channels are more calcium sensitive than BK channels in other tissues where the $\beta 1$ subunit is not expressed²⁵.

Although the functional role of the BK $\beta 1$ subunit in native tissues is unknown, the assumption is that the $\beta 1$ subunit may increase calcium sensitivity sufficiently for regulation of smooth muscle membrane properties. However, evidence indicates that BK channels are localized very close to calcium release sites, such that they would be exposed to effectively high calcium concentrations ($>10 \mu\text{M}$ calcium)²⁶ during calcium release. Thus, it is not known whether the increased BK channel calcium sensitivity conferred by the $\beta 1$ subunit contributes to normal smooth muscle function.

Our hypothesis was that the $\beta 1$ subunit has a central molecular role in conferring specificity to the translation of local calcium signals to long-distance electrical events (that is, membrane hyperpolarization) so as to couple calcium to a decrease in vascular tone and blood pressure. We have used gene targeting to eliminate $\beta 1$ subunit expression in mice and evaluate its contribution to vascular smooth muscle function. We examined the roles of the $\beta 1$ subunit in determining the Ca^{2+} sensitivity of BK channels, and the coupling of Ca^{2+} sparks to BK channel activity in smooth muscle cells from cerebral arteries. Furthermore, the $\beta 1$ knockout mice allowed us to evaluate the functional role of the $\beta 1$ subunit in regulating arterial tone and systemic blood pressure in awake, resting animals.

Our results indicate that the calcium sensitivity of the BK channel is uniquely matched by β -subunit association to a given calcium signalling modality to modulate the physiology of arteries.

Generation of $\beta 1$ knockout mice

To create a null allele of the $\beta 1$ locus, the gene-targeting vector was constructed to delete the first coding exon (exon 2) of the gene²⁷. Exon 2 encodes the amino-terminal coding sequence including the

first transmembrane domain of the $\beta 1$ protein. The targeting vector was designed to insert a β -galactosidase reporter in translational frame with the $\beta 1$ subunit translation initiation site (Fig. 1a), and thus report transcription from the $\beta 1$ gene promoter. Transfection of the targeting construct in embryonic stem cells generated three independent clones that specifically targeted the mouse $\beta 1$ gene. Each of the embryonic stem cell clones was implanted into blastocysts and produced germline transmitting mice. Genomic Southern analysis using DNA fragments to hybridize outside the left and right arms of the targeting vector confirmed the recombination (Fig. 1b), seen as a polymorphism that shifts the bands to a predictable, larger size (Fig. 1). To confirm the disruption of the $\beta 1$ gene, an antisense RNA probe encompassing the N-terminal coding region was used as a probe against RNA extracted from $\beta 1$ knockout and control stomach tissue. RNase protection (Fig. 1c)

reveals a protected $\beta 1$ probe hybridizing to control but not knock-out mice RNA, confirming the lack of $\beta 1$ expression in the knockout mice.

$\beta 1$ messenger RNA is enriched in smooth muscle

The $\beta 1$ subunit mRNA has been detected in many tissues containing smooth muscle²⁷. The lacZ gene was targeted to the $\beta 1$ locus to permit the examination of the cell types that normally express the $\beta 1$ subunit. Figure 2 shows lacZ staining in isolated cerebral arteries used in this study as well as other smooth muscle tissues. Within certain tissues, expression is restricted to arterial smooth muscle. Examples are brain and heart (Fig. 2c, g), where $\beta 1$ expression is largely undetected except in the vasculature, such as the cerebral arteries (Fig. 2b), aorta and coronary arteries (not shown). This is consistent with *in situ* hybridizations of $\beta 1$ RNA, where expression

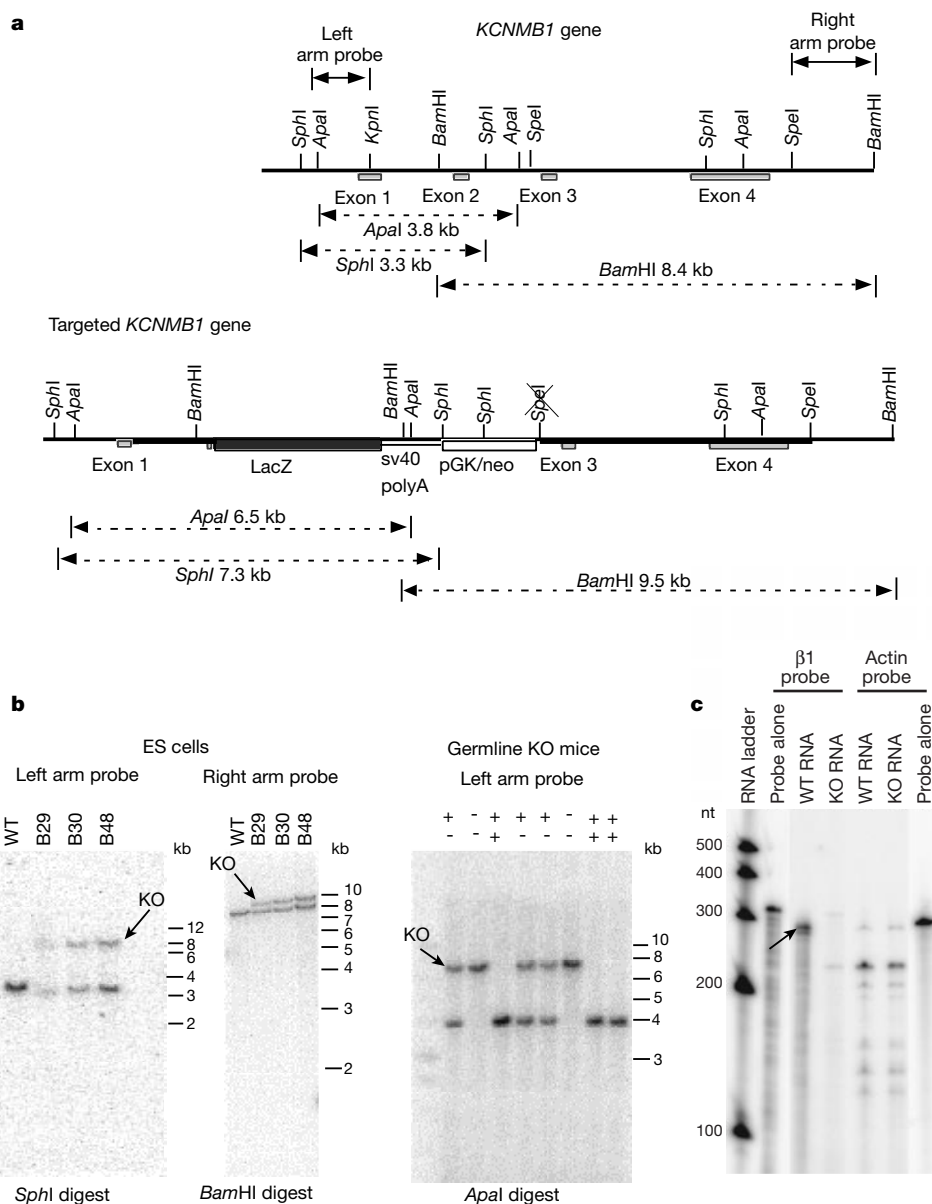


Figure 1 Generation of $\beta 1$ gene knockout mice. **a**, Restriction map of the $\beta 1$ gene locus (top) and the targeted $\beta 1$ gene (bottom). Thick line in the targeted locus map designates the regions used for the targeting vector. **b**, Southern and RNA analysis of the gene knockouts. Left, Southern analysis of ES cells using the probes designated in **a**. Lanes contain genomic DNA from control ES cells (WT) or the $\beta 1$ -targeted ES clones (B29, B30 and B48). Middle, genomic Southern analysis of $\beta 1$ knockout siblings derived from clone

B30 heterozygous parent matings. Lanes are labelled with the genotypes. **c**, RNase protection assay of $\beta 1$ expression from control (WT) and $\beta 1$ knockout (KO) RNA. The antisense $\beta 1$ probe hybridizes to a specific 290-nucleotide (nt) product in WT RNA (arrow) but not to the $\beta 1$ knockout RNA. A mouse β -actin RNA probe hybridizes to a specific 245-nt product in both WT and KO RNA samples.

was enriched in aortic smooth muscle but absent from brain¹⁹. Expression of $\beta 1$ was also observed in other tissues, including smooth muscle of the bladder, trachea, bronchi and the digestive tract (trachea, stomach and colon staining are shown in Fig. 2d–f, respectively).

$\beta 1$ knockout BK channels have reduced Ca^{2+} sensitivity

The fact that the $\beta 1$ subunit is restricted in expression to smooth muscle cells suggested a unique role of the $\beta 1$ subunit in Ca^{2+} signalling in smooth muscle. To confirm that the arterial smooth muscle from the knockout mice lacked a functional $\beta 1$ subunit, the sensitivity of BK channels to calcium and the BK/ $\beta 1$ channel agonist DHS-1 were examined in inside-out patches from freshly isolated cerebral artery myocytes at physiological membrane potentials (-40 mV) for pressurized arteries^{4,5}. Figure 3 compares single BK channel recordings of excised patches from knockout and control arterial smooth muscle. BK channels were exposed to cytoplasmic calcium concentrations of 3 and 10 μM , concentrations within the range predicted for calcium-spark-evoked activation of BK channels in arterial smooth muscle²⁶. At -40 mV, control BK channels have a significant open probability when activated by 3 μM ($P_o = 0.47 \pm 0.2$) and by 10 μM Ca^{2+} ($P_o = 0.60 \pm 0.2$). In contrast, BK channels from KO cells had an open probability at least 100-fold lower at -40 mV ($P_o = 0.003 \pm 0.002$, 0.007 ± 0.002 for 3 and 10 μM calcium, respectively). Increasing the membrane voltage to $+40$ mV increases the open probability in both control and KO channels, but the control BK channels still have significantly higher open probability than the BK channels in the knockout cells (control, $P_o = 0.89 \pm 0.03$, 0.78 ± 0.073 ; knockout, $P_o = 0.22 \pm 0.07$, 0.53 ± 0.06 , for 3 and 10 μM calcium, respectively).

The agonist DHS-1 significantly increases the open probability of the channel when associated with a $\beta 1$ subunit²⁰ and provides a pharmacological probe for the presence of BK $\alpha/\beta 1$ -assembled

subunits. Figure 4a shows that application of DHS-1 to the bath causes a significant increase in open probability in control mice ($P_{o\text{control}} = 0.11 \pm 0.1$ versus $P_{o\text{control-DHS1}} = 0.25 \pm 0.1$; $P < 0.05$, $n = 3$, -40 mV). DHS-1, however, had no effect on the BK channels from $\beta 1$ knockout mice ($P_{o\text{KO}} = 0.005 \pm 0.001$ versus $P_{o\text{KO-DHS1}} = 0.007 \pm 0.001$; $P > 0.4$, $n = 3$, -40 mV). Neither BK channel density, as assayed from the average number of channels detected in patches (Fig. 4b), nor the BK channel unitary conductance (γ), was affected in the knockout ($\gamma_{\text{KO}} = 209 \pm 14$ pS, $\gamma_{\text{control}} = 215 \pm 1.1$ pS, $n = 3$ patches each, from -90 to $+60$ mV in symmetric 140 mM K, $P > 0.4$). The decreased open probability and insensitivity to DHS-1 support the conclusion that BK channels in normal cerebrovascular smooth muscle consist primarily of BK $\alpha/\beta 1$ subunits, whereas the $\beta 1$ knockout mice contain BK channels lacking an associated $\beta 1$ subunit. This is consistent with the LacZ staining demonstrating $\beta 1$ subunit expression in cerebral artery smooth muscle (Fig. 2).

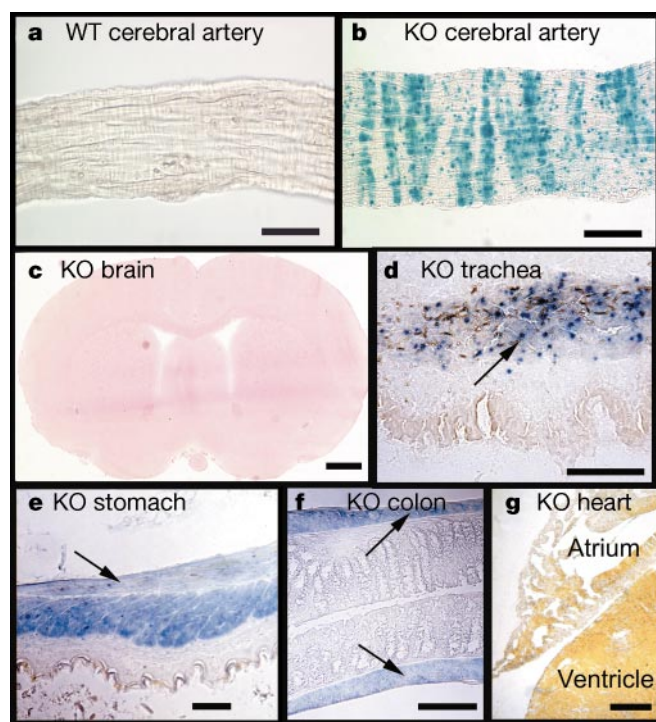


Figure 2 Detection of lacZ gene expression from $\beta 1$ gene-targeted mice. Staining of whole-mount tissue (a, b) and frozen sections (c–f) from knockout (KO) and control (WT) mice. Blue staining (arrows) labels regions of $\beta 1$ expression. Scale bars: a, b, d, e, 50 μm ; c, 1 mm; f, 200 μm ; g, 500 μm . Tissues are counterstained with Orange-G in d–g and with Eosin Y in c.

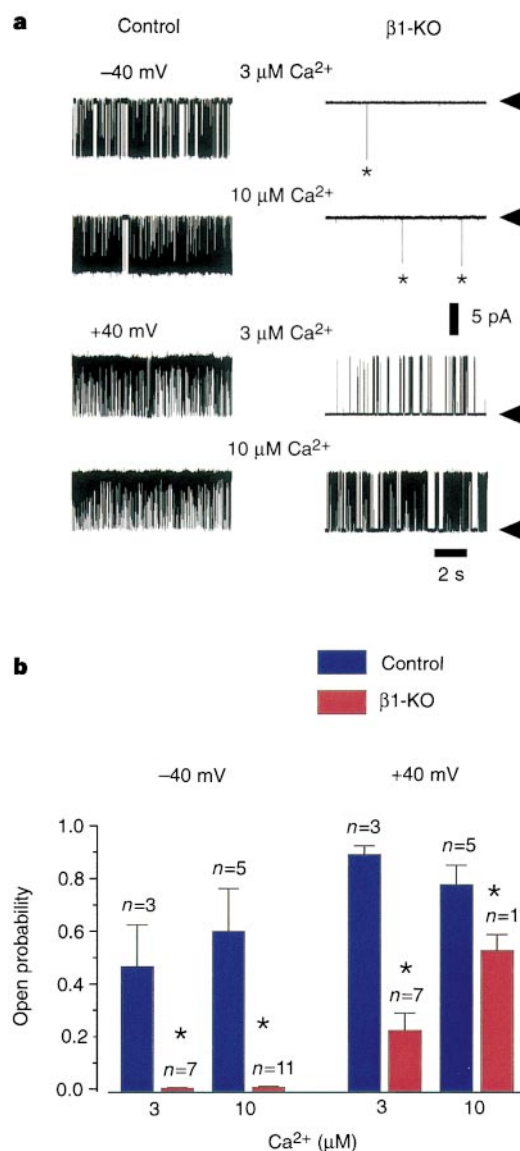


Figure 3 Ca^{2+} and voltage dependence of BK channels in cerebral artery myocytes from control and $\beta 1$ -KO animals. a, Single channel recordings in inside-out patches held at -40 and $+40$ mV, and with 3 and 10 μM free Ca^{2+} . Arrows, closed state. Asterisks, brief channel openings in $\beta 1$ -KO patches. b, BK channel open probability in control and $\beta 1$ -KO animals at two different voltages (-40 and $+40$ mV) and two Ca^{2+} concentrations (3 and 10 μM). Asterisks denote statistically significant difference from controls.

β 1 subunit couples calcium sparks to BK channel activation

The open probability of BK channels is low at physiological membrane potentials (-40 mV) and average arterial smooth muscle calcium of pressurized cerebral arteries⁴. A calcium spark can elevate the P_o of nearby BK channels 10^4 to 10^6 -fold to produce a significant transient membrane potential hyperpolarization (20 mV)^{2,26}. To explore the role of the β 1 subunit in the communication of calcium sparks to BK channels, we measured calcium sparks and whole-cell potassium currents simultaneously in isolated cerebral artery myocytes²⁸. Figure 5a illustrates the life cycle of a spark from a β 1 knockout mouse, peaking in around 20 ms and then decaying over 200 ms, very similar to control myocytes²⁶. Figure 5b shows a representative simultaneous recording of whole-cell current (in blue) and sparks (in red and green). Transient BK current amplitude increases with Ca^{2+} spark amplitude in both control and knockout cells (Fig. 5c). However, the mean transient BK current amplitude was around one-sixth of control amplitude for a given Ca^{2+} spark amplitude. Furthermore, there was a striking difference in the ability of Ca^{2+} sparks to activate transient BK currents. In the control, essentially every Ca^{2+} spark evoked a transient BK current at -40 mV. In contrast, in the knockout 35% of the sparks failed to evoke a detectable BK current (Fig. 5d). BK channel density, Ca^{2+} spark amplitude (control = 1.65 ± 0.03 ($n = 94$ from 6 cells), knockout = 1.77 ± 0.05 ($n = 71$ from 7 cells)), and spark frequency (control = 1.41 ± 0.2 Hz (6 cells), and knockout = 1.35 ± 0.3 (7 cells)) were unaltered in the knockout. These results indicate that the overall BK channel activity during Ca^{2+} sparks is reduced at least 12-fold in the knockout, consistent with the diminished Ca^{2+} sensitivity of BK channels (Figs 3 and 4).

Lack of β 1 elevates arterial tone and blood pressure

BK channels lacking the β 1 subunit are present and functional in arteries from knockout mice, albeit less sensitive to activating calcium. To understand the physiological consequences of the reduced coupling between calcium sparks and BK channel activation

in the β 1 knockouts, we evaluated the effect of pressure on arterial diameter. Elevation of intravascular pressure constricts small arteries, including cerebral arteries^{4,29–31}. Cerebral arteries that lack the β 1 subunit are significantly more constricted at a given pressure than are control arteries (Fig. 6a–c). These results indicate that the lack of the β 1 subunit leads to an elevation in arterial tone.

The contribution of the β 1 subunit to the regulation of arterial tone can be evaluated by examining the effects of the BK inhibitor iberiotoxin (IBTX) on arterial diameter. IBTX caused a 74% increase in arterial tone in the control (Fig. 6d, f). In contrast, IBTX did not affect knockout cerebral arteries (Fig. 6e, f). These results indicate that BK channels lacking the β 1 subunit are unable to contribute to the regulation of arterial tone.

Unless systemic physiological control mechanisms can compensate for the increased arterial tone, arterial blood pressure should be elevated in mice lacking the BK β 1 subunit. The mean arterial blood pressure of the knockout mice was indeed elevated and comparable to transgenic mice with compromised endothelial function³². The knockout lines were bred from a mixture of 129svj mice, used to derive the embryonic stem cells, and C57BL mice, used as the initial breeder mates. We chose to compare the mean arterial pressure of our knockout mice to the mean arterial pressure of 129svj control mice to generate the most stringent test for evidence of elevated blood pressure, as the 129svj mice have a higher mean arterial blood pressure than any other strain³³. The knockout mice exhibited an increase in mean blood pressure, even over the pure 129svj mice (Fig. 7a). These data indicate that the increased arterial tone leads to an elevation in blood pressure.

In humans, long-standing essential hypertension induces significant left ventricular hypertrophy and leads to heart enlargement³⁴.

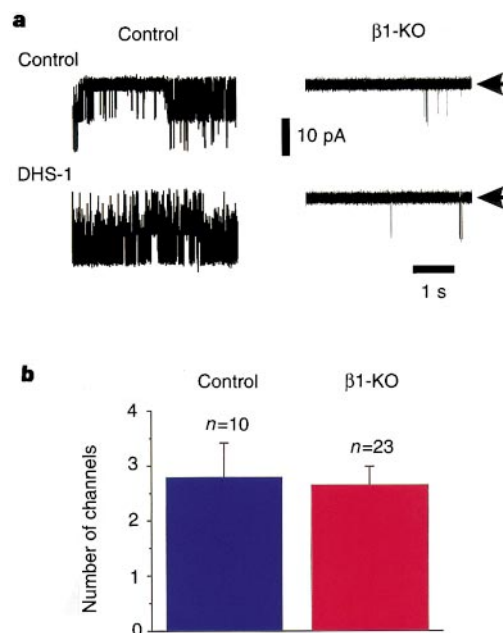


Figure 4 DHS-1 sensitivity and density of BK channels in cerebral arteries myocytes from control and β 1-KO animals. **a**, Single channel recordings at -40 mV and $10 \mu\text{M}$ Ca^{2+} , before and after the addition of 100 nM DHS-1. **b**, Average number of channels per patch excised from control and β 1-KO myocytes (10 and 23 patches, respectively). No significant difference was found between control and β 1-KO patches. Pipette resistances were 8 – 11 M Ω .

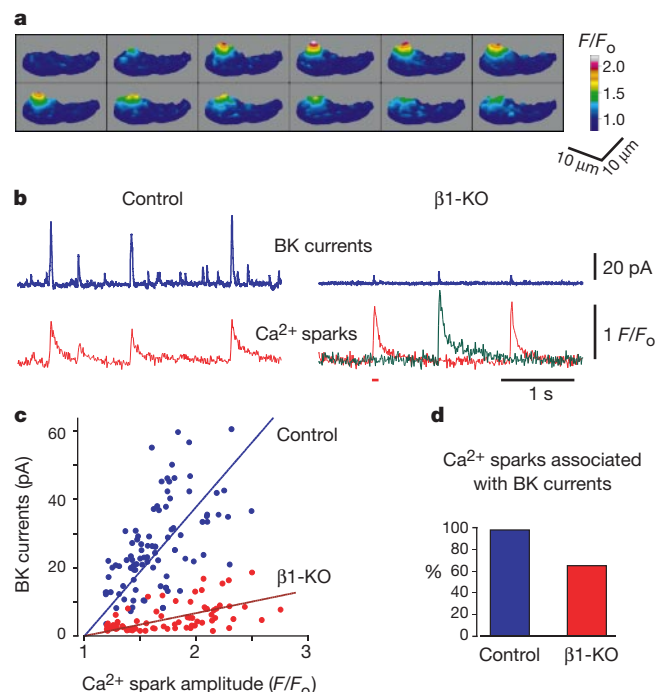


Figure 5 Decreased coupling of calcium sparks to BK channels in β 1-KO myocytes. **a**, Consecutive pseudocolor three-dimensional images of a β 1-KO cell obtained every 8.33 ms. **b**, Simultaneous BK current (blue) and Ca^{2+} spark measurements (fractional fluorescence, F/F_0) from a control cell with one spark site and a β 1-KO cell with two spark sites (red and green) (-40 mV). Red bar indicates the segment of the red trace illustrated in **a**. **c**, Relationships between BK current and Ca^{2+} spark amplitudes in control cells (blue, 94 sparks, 6 cells) and β 1-KO cells (red, 71 sparks, 7 cells). Lines represent linear regression fit (slope_{control} = 37.8 ± 1.6 versus slope β 1-KO = 6.7 ± 0.5 , $P < 0.001$). **d**, Percentage of Ca^{2+} sparks causing transient BK currents in control (blue) and β 1-KO (red) cells.

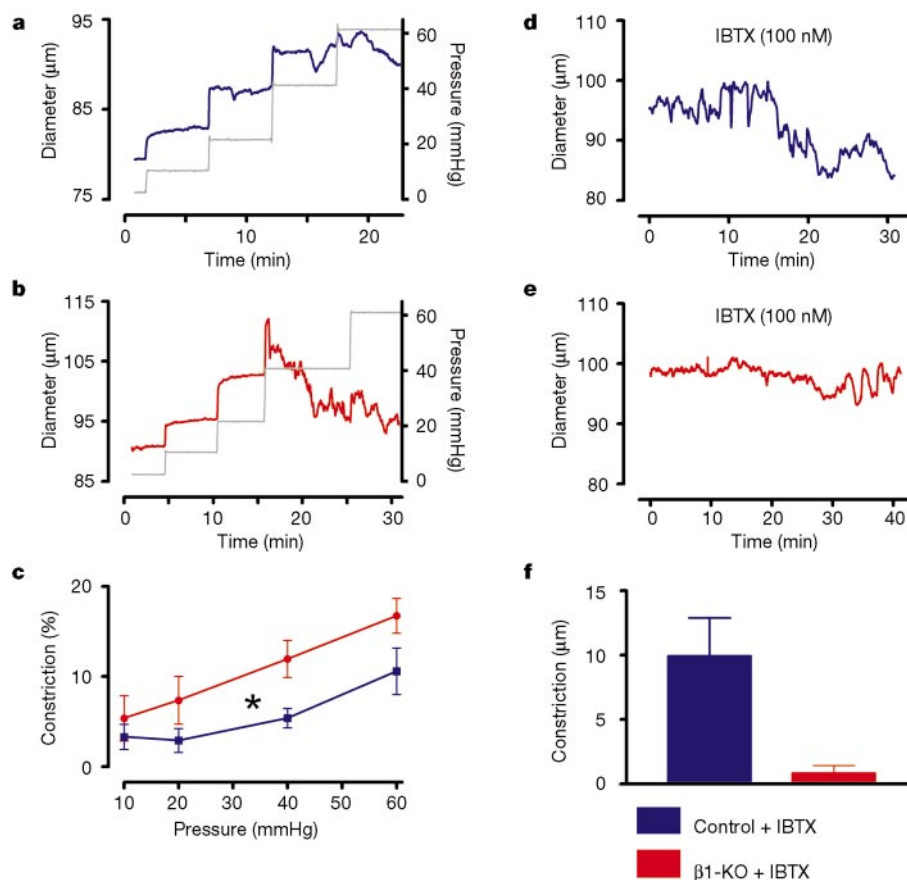


Figure 6 $\beta 1$ -KO cerebral arteries are more constricted to pressure. **a**, Control, blue; **b**, $\beta 1$ -KO, red; **c**, summary, $P < 0.05$. They are also insensitive to the BK channel

blocker iberiotoxin (**d**, control, blue; **e**, $\beta 1$ -KO, red; **f**, summary). Grey lines indicate pressure levels in **a** and **b**.

To measure heart enlargement, we normalized the heart weight to animal body weight. Knockout mice had larger heart-to-body weight ratios than 129svj mice (Fig. 7b). Furthermore, at the electron microscopy level, there were no obvious differences in ultrastructure of the cardiomyocytes between the 129svj control hearts and the $\beta 1$ knockout hearts (Fig. 7c). Myofibrillar organization, mitochondrial prevalence and structure, and glycogen content appeared similar in control and knockout mice. There was no evidence of necrosis or fibrosis. As heart muscle ultrastructure appeared normal, these findings are consistent with heart enlargement caused by uncomplicated essential hypertension.

Discussion

The BK channel is the only member of the voltage-dependent potassium channel family that is activated by both voltage and calcium. This makes it particularly suited to integrate calcium and voltage signals to modulate membrane excitability in a variety of cell types. But the tissues that express BK channels have diverse functions with considerable differences in excitability and Ca^{2+} signaling. Our results support the concept that the $\beta 1$ subunit is required specifically to tune BK channel properties to the needs of an arterial smooth muscle cell. The increased sensitivity to calcium conferred by the $\beta 1$ subunit is required for the BK channel to translate calcium sparks to membrane potential hyperpolarization. The decreased coupling of Ca^{2+} to channel activity extrapolated well to the functional defects observed in the intact artery and whole animal. Moreover, end organ pathology observed in chronic hypertension, such as myocardial hypertrophy, was also observed in the BK $\beta 1$ knockout mice. The BK $\beta 1$ knockout mouse therefore presents a unique model, wherein a clearly defined molecular defect could be used to study the secondary effects of hypertension. Moreover, the

$\beta 1$ subunit gene should provide a candidate genetic locus for human hypertension.

Not only was BK channel activity reduced when the $\beta 1$ subunit was absent, as detected by the reduced size of transient BK currents, but many calcium sparks failed to cause detectable currents. The most direct explanation is that the calcium sensitivity of the BK channel lacking $\beta 1$ is reduced sufficiently so that the channel open probability during a spark is too low to cause a detectable current (one single BK channel amplitude at -40 mV, or 2 pA). The effect is to uncouple BK channel activation from calcium spark signals. Further evidence of uncoupling was the lack of effect of iberiotoxin on arterial tone in $\beta 1$ knockout arteries. It has been estimated that BK channels in the presence of $\beta 1$ subunits appose close enough to calcium release sites to detect 10–100 μM effective calcium concentrations²⁶. In 10 μM calcium, cloned vascular smooth muscle BK channels lacking a $\beta 1$ subunit have a voltage for half activation of around +30 mV³⁵, and have a very low open probability at -40 mV, the membrane potential of smooth muscle in intact pressurized arteries. In contrast, the same channels containing the $\beta 1$ subunit have activation voltages 70 mV more negative³⁵, at approximately the resting membrane potential for smooth muscle cells (-40 mV). Thus, the reduced apparent calcium sensitivity can explain the difference in spark/BK current coupling between the knockout mice and their controls. However, this does not exclude the possibility that the $\beta 1$ subunit may have other effects on BK channel properties. For example, $\beta 1$ subunits could be required for subcellular localization of BK channels to calcium release sites or modifying BK channel phosphorylation-dependent regulation.

The BK channel regulates arterial diameter and mediates the response to a number of smooth muscle relaxants including nitric

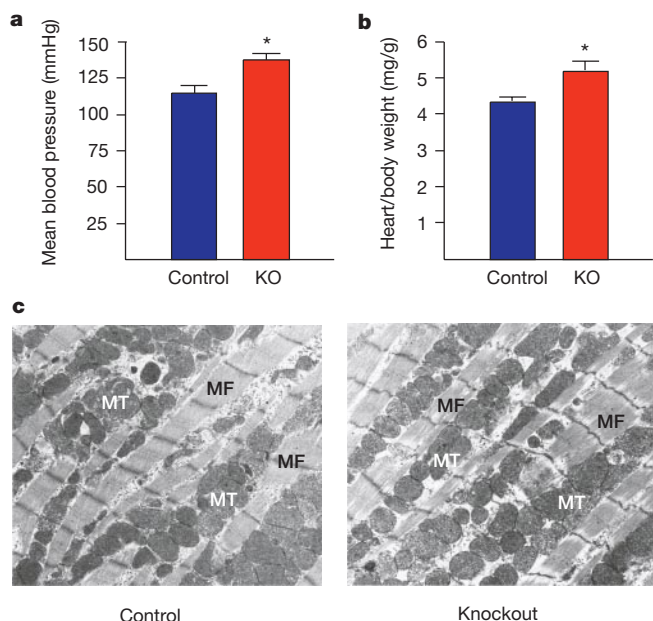


Figure 7 $\beta 1$ -KO mice show symptoms of hypertension. **a**, Mean blood pressure of $\beta 1$ -KO and control 129svj mice. Control average blood pressure is 114 ± 6.0 mm Hg ($n = 6$), KO average blood pressure is 134 ± 5.1 mm Hg ($n = 6$), $P = 0.029$. **b**, Heart/body weight measurements comparing $\beta 1$ -KO and control mice. KO is 5.0 ± 0.22 mg/g ($n = 11$), controls are 4.35 ± 0.15 mg/g ($n = 8$), $P = 0.037$. **c**, Electron microscopy of heart tissue from $\beta 1$ -KO and control mice. Examples of myofibril (MF) and mitochondria (MT) are labelled.

oxide^{36–38}. Our results support the concept that the calcium sensitivity of the BK channel is fine tuned to respond to calcium signals unique to the physiology of a given cell type. In the case of arterial smooth muscle, the $\beta 1$ subunit is essential for the effective coupling of calcium sparks to BK channels, thereby enabling BK channel regulation of arterial smooth muscle tone. In different tissues, other β -subunit family members may serve similar roles. Furthermore, we propose that other agents, such cyclic AMP- and cGMP-mediated vasodilators, which modulate the calcium sensitivity of the BK channels, can dynamically match calcium signals to BK channel activation to regulate cell function. □

Methods

$\beta 1$ knockout mice

To generate a left arm fragment, we amplified a 129svj/lambda FixII genomic clone of the $\beta 1$ gene with a 3' primer that overlapped the $\beta 1$ seventh codon (methionine) and created a flanking *KpnI* site. The 5' primer introduced a *NotI* site that annealed 1.6 kilobases (kb) upstream (in the first exon²⁷). The left arm PCR product was ligated in a three-way ligation with a *KpnI/SalI* fragment containing the lacZ/SV40 polyA sequence and a *NotI/XhoI* cut pPNT knockout vector³⁹. The right arm was a 5.4-kb *Spe* genomic fragment (Fig. 1) cloned into the *XbaI* site of the pPNT knockout vector³⁹. The construct was electroporated into 129svj ES cells. We screened candidate clones by genomic Southern analysis of the short arm (Fig. 1) and confirmed them by Southern analysis of the long arm. Targeted ES cells were used to generate germline carrying mice⁴⁰. Germline siblings (mixed 129svj/C57BL background) were mated to produce homozygous stocks. 129svj mice were used as controls for blood pressure and heart/body weight measurements and either 129svj or F_1 progeny from a 129svj/C57BL cross were used for other experiments. Animals used were males between 3 and 8 months of age for blood pressure and heart size measurements, or either males or females of 3–8-month age range for cell physiology experiments.

DNA and RNA analysis

Genomic Southern analysis was as described⁴¹. To make an antisense $\beta 1$ RNA, nucleotides 1–290 of the coding portion of the mouse $\beta 1$ complementary DNA were subcloned into the vector pOX¹¹ and transcribed using the T7 RNA polymerase promoter. RNase protection assays were conducted using the Ambion RPAII kit according to the manufacturer's protocols (Ambion). Hybridization was conducted with 2 μ g of poly(A) purified stomach RNA to detect $\beta 1$ or 200 ng total stomach RNA to detect actin.

LacZ staining of tissue and electron microscopy

Frozen sections (16 μ m) were fixed in 0.2% glutaraldehyde in PBS, washed three times in

PBS and then stained for β -galactosidase activity as described⁴². The tissues were counterstained with Eosin Y or Orange G. Electron microscopy of heart tissue was done as described⁴³.

Tissue preparation

Knockout or control mice (129/sv or 129/svXB6) of either sex were killed by exsanguination while under deep pentobarbital anaesthesia (intraperitoneal; 150 mg kg^{−1} body weight). Cerebral arteries were rapidly dissected while the brain was submerged in cold (4 °C) oxygenated (95% O₂/5% CO₂) physiological saline solution (PSS) of the following composition (in mmol l^{−1}): 118.5 NaCl, 4.7 KCl, 24 NaHCO₃, 1.18 KH₂PO₄, 2.5 CaCl₂, 1.2 MgCl₂, 0.023 EDTA, 11 glucose.

Diameter measurements

We measured the diameter of middle cerebral artery segments by video edge detection, in oxygenated PSS at 37 °C and pH 7.4 (refs 1, 2, 4, 5, 44, 45). The diameters (at 10 mm Hg) from control and $\beta 1$ -knockout arteries were 92.0 ± 2.5 μ m ($n = 11$) and 84.5 ± 2.7 μ m, $n = 12$ respectively. Pressure-induced constrictions are expressed as a per cent decrease of the diameters in Ca²⁺-free PSS containing 1 μ mol l^{−1} nisoldipine.

Electrical and optical measurements

Single smooth muscle cells were isolated from cerebral arteries by digestion in 0.3 mg ml^{−1} papain and 1 mg ml^{−1} dithioerythritol for 10 min, and then transferred for a second digestion in 1 mg ml^{−1} collagenase (type F and type H in a 70%–30% mixture) for 7 min^{2,26,45}. Membrane current and calcium sparks were measured simultaneously in myocytes loaded with the Ca²⁺ indicator fluo-4, using perforated patch, whole-cell configuration and a laser (two-dimensional) scanning confocal microscope^{26,46}. The bathing solution was (in mM) 134 NaCl, 6 KCl, 1 MgCl₂, 2 CaCl₂, 10 glucose, 10 HEPES pH 7.4, with 5 μ M wortmannin to minimize contraction. The pipette solution contained (in mM): 110 K-aspartate, 30 KCl, 10 NaCl, 1 MgCl₂, 10 HEPES (pH 7.2) and 250 mg ml^{−1} amphotericin B. Single BK channel currents were recorded in inside-out patches, exposed to calibrated Ca²⁺ buffered solutions (in mM): 140 KCl, 10 HEPES (pH 7.2), 1 Mg²⁺, 5 HEDTA, and 3 or 10 μ M Ca²⁺ free (adjusted with Ca²⁺ electrodes). Pipette solution contained (in mM): 140 KCl, 10 HEPES (pH 7.2), 1 Mg²⁺, 5 HEDTA and 10 μ M Ca²⁺. Single-channel currents were recorded over 1–10 min at steady potentials of −40 and +40 mV at room temperature. Currents were filtered at 1 kHz and digitized at 4 kHz.

Blood pressure analysis and heart/body weight ratios

Mean arterial blood pressures were obtained using arterial catheters surgically inserted into the left carotid artery as described³³. Following surgery, animals were allowed to recover for 24 h, and mean arterial blood pressure was measured every 8 s for 30 min. The average mean arterial blood pressure values obtained during the sampling period was calculated for each animal. All measurements were performed with the animal resting quietly. To obtain heart weights, isolated beating hearts were dissected free and allowed to expel the blood volume in saline solution. Hearts were transferred to 1 M KCl solution for 15–30 s, washed in saline, blotted free of excess solution and weighed. For blood pressure measurements the knockout animals had a mean age of 17.7 ± 2.2 weeks, and control animals were 20.2 ± 3.5 weeks. For heart/body weights, the animals were matched well for body weights (knockout, 32.1 ± 1.6 versus controls, 31.7 ± 1.2 g) although not ideally for age (knockout 15.6 ± 0.9 weeks). Our data and others show no correlation between increasing age and heart/body weights in normal animals^{47,48}.

Data analysis

Image and BK current analysis was performed using a custom written analysis program using Interactive Data Language software (Research Systems Inc). Baseline fluorescence (F_0) was determined by averaging 30 images with no activity. Fractional fluorescence (F/F_0) versus time traces were obtained by averaging F/F_0 from a box region of 2.2×2.2 μ m centred in the active area of interest (Ca²⁺ sparks sites). Transient BK currents caused by Ca²⁺ sparks with amplitudes bigger than a single BK channel opening (2 pA, at −40 mV) were considered for further analysis. Number of channels (N) $\times$ open probability (P_o) of single channel currents was calculated from all points histograms and divided by the number of channels present in the patch to obtain P_o values.

Statistical analysis

Results are expressed as means $\pm$ s.e.m. where applicable. Comparisons between control and knockout data were done with the unpaired two-tailed student's *t*-test.

Received 6 July; accepted 30 August 2000.

1. Knot, H. J., Zimmermann, P. A. & Nelson, M. T. Extracellular K⁺-induced hyperpolarizations and dilations of rat coronary and cerebral arteries involve inward rectifier K(+) channels. *J. Physiol. (Lond.)* **492**, 419–430 (1996).
2. Jaggar, J. H., Porter, V. A., Lederer, W. J. & Nelson, M. T. Calcium sparks in smooth muscle. *Am. J. Physiol. Cell Physiol.* **278**, C235–C256 (2000).
3. Galvez, A. et al. Purification and characterization of a unique, potent, peptidyl probe for the high conductance calcium-activated potassium channel from venom of the scorpion *Buthus tamulus*. *J. Biol. Chem.* **265**, 11083–11090 (1990).
4. Knot, H. J., Standen, N. B. & Nelson, M. T. Ryanodine receptors regulate arterial diameter and wall [Ca²⁺] in cerebral arteries of rat via Ca²⁺-dependent K⁺ channels. *J. Physiol. (Lond.)* **508**, 211–221 (1998).
5. Brayden, J. E. & Nelson, M. T. Regulation of arterial tone by activation of calcium-dependent potassium channels. *Science* **256**, 532–535 (1992).

6. Nelson, M. T. & Quayle, J. M. Physiological roles and properties of potassium channels in arterial smooth muscle. *Am. J. Physiol.* **268**, C799–822 (1995).
7. Sah, P. Ca^{2+} -activated K^+ currents in neurones: types, physiological roles and modulation. *Trends Neurosci.* **19**, 150–154 (1996).
8. Vergara, C., Latorre, R., Marrion, N. V. & Adelman, J. P. Calcium-activated potassium channels. *Curr. Opin. Neurobiol.* **8**, 321–329 (1998).
9. Kaczorowski, G. J., Knaus, H. G., Leonard, R. J., McManus, O. B. & Garcia, M. L. High-conductance calcium-activated potassium channels; structure, pharmacology, and function. *J. Bioenerg. Biomembr.* **28**, 255–267 (1996).
10. Latorre, R. in *Molecular Workings of Large Conductance (Maxi) Ca^{2+} -Activated K^+ Channels* 79–102 (Academic, San Diego, 1994).
11. Brenner, R., Jegla, T. J., Wickenden, A., Liu, Y. & Aldrich, R. W. Cloning and functional characterization of novel large conductance calcium-activated potassium channel beta subunits, hKCNMB3 and hKCNMB4. *J. Biol. Chem.* **275**, 6453–6461 (2000).
12. Behrens, R. *et al.* hKCNMB3 and hKCNMB4, cloning and characterization of two members of the large-conductance calcium-activated potassium channel beta subunit family. *FEBS Lett.* **474**, 99–106 (2000).
13. Knaus, H. G. *et al.* Primary sequence and immunological characterization of beta-subunit of high conductance Ca^{2+} -activated K^+ channel from smooth muscle. *J. Biol. Chem.* **269**, 17274–17278 (1994).
14. Meera, P., Wallner, M. & Toro, L. A neuronal beta subunit (KCNMB4) makes the large conductance, voltage- and Ca^{2+} -activated K^+ channel resistant to charybdotoxin and iberiotoxin. *Proc. Natl Acad. Sci. USA* **97**, 5562–5567 (2000).
15. Wallner, M., Meera, P. & Toro, L. Molecular basis of fast inactivation in voltage and Ca^{2+} -activated K^+ channels: a transmembrane beta-subunit homolog. *Proc. Natl Acad. Sci. USA* **96**, 4137–4142 (1999).
16. Xia, X. M., Ding, J. P. & Lingle, C. J. Molecular basis for the inactivation of Ca^{2+} - and voltage-dependent BK channels in adrenal chromaffin cells and rat insulinoma tumor cells. *J. Neurosci.* **19**, 5255–5264 (1999).
17. Xia, X. M., Ding, J. P., Zeng, X. H., Duan, K. L. & Lingle, C. J. Rectification and rapid activation at low Ca^{2+} of Ca^{2+} -activated, voltage-dependent BK currents: consequences of rapid inactivation by a novel beta subunit. *J. Neurosci.* **20**, 4890–4903 (2000).
18. Riaz, M. A. *et al.* Identification of a putative regulatory subunit of a calcium-activated potassium channel in the dup(3q) syndrome region and a related sequence on 22q11.2. *Genomics* **62**, 90–94 (1999).
19. Weiger, T. M. *et al.* A novel nervous system beta subunit that downregulates human large conductance calcium-dependent potassium channels. *J. Neurosci.* **20**, 3563–3570 (2000).
20. McManus, O. B. *et al.* Functional role of the beta subunit of high conductance calcium-activated potassium channels. *Neuron* **14**, 645–650 (1995).
21. Knaus, H. G., Eberhart, A., Kaczorowski, G. J. & Garcia, M. L. Covalent attachment of charybdotoxin to the beta-subunit of the high conductance Ca^{2+} -activated K^+ channel. Identification of the site of incorporation and implications for channel topology. *J. Biol. Chem.* **269**, 23336–23341 (1994).
22. Dworetzky, S. I. *et al.* Phenotypic alteration of a human BK (hSlo) channel by hSlobeta subunit coexpression: changes in blocker sensitivity, activation/relaxation and inactivation kinetics, and protein kinase A modulation. *J. Neurosci.* **16**, 4543–4550 (1996).
23. Cox, D. H. & Aldrich, R. W. Regulation of the large conductance Ca^{2+} -activated K^+ channel by its β subunit. *J. Gen. Physiol.* **116**, 411–432 (2000).
24. Meera, P., Wallner, M., Jiang, Z. & Toro, L. A calcium switch for the functional coupling between alpha (hslo) and beta subunits (K_{VCa} beta) of maxi K channels. *FEBS Lett.* **382**, 84–88 (1996).
25. Tanaka, Y., Meera, P., Song, M., Knaus, H. G. & Toro, L. Molecular constituents of maxi K_{Ca} channels in human coronary smooth muscle: predominant alpha + beta subunit complexes. *J. Physiol. (Lond.)* **502**, 545–557 (1997).
26. Perez, G. J., Bonev, A. D., Patlak, J. B. & Nelson, M. T. Functional coupling of ryanodine receptors to K_{Ca} channels in smooth muscle cells from rat cerebral arteries. *J. Gen. Physiol.* **113**, 229–238 (1999).
27. Jiang, Z., Wallner, M., Meera, P. & Toro, L. Human and rodent MaxiK channel beta-subunit genes: cloning and characterization. *Genomics* **55**, 57–67 (1999).
28. Poolos, N. P. & Johnston, D. Calcium-activated potassium conductances contribute to action potential repolarization at the soma but not the dendrites of hippocampal CA1 pyramidal neurons. *J. Neurosci.* **19**, 5205–5212 (1999).
29. Bayliss, W. M. On the local reactions of the arterial wall to changes in internal pressure. *J. Physiol. (Lond.)* **28**, 220–231 (1902).
30. Nelson, M. T. *et al.* Relaxation of arterial smooth muscle by calcium sparks. *Science* **270**, 633–637 (1995).
31. Meininger, G. A. & Davis, M. J. Cellular mechanisms involved in the vascular myogenic response. *Am. J. Physiol.* **263**, H647–659 (1992).
32. Shesely, E. G. *et al.* Elevated blood pressures in mice lacking endothelial nitric oxide synthase. *Proc. Natl Acad. Sci. USA* **93**, 13176–13181 (1996).
33. Desai, K. H. *et al.* Cardiovascular indexes in the mouse at rest and with exercise: new tools to study models of cardiac disease. *Am. J. Physiol.* **272**, H1053–1061 (1997).
34. Wilson, P. W. From hypertension to heart failure: what have we learned? *Clin. Cardio.* **22** (suppl. 5), V1–10 (1999).
35. McCobb, D. P. *et al.* A human calcium-activated potassium channel gene expressed in vascular smooth muscle. *Am. J. Physiol.* **269**, H767–777 (1995).
36. Zhou, X. B., Schlossmann, J., Hofmann, F., Ruth, P. & Korth, M. Regulation of stably expressed and native BK channels from human myometrium by a cGMP- and cAMP-dependent protein kinase. *Pflügers Arch.* **436**, 725–734 (1998).
37. Dong, H., Waldron, G. J., Cole, W. C. & Triggie, C. R. Roles of calcium-activated and voltage-gated delayed rectifier potassium channels in endothelium-dependent vasorelaxation of the rabbit middle cerebral artery. *Br. J. Pharmacol.* **123**, 821–832 (1998).
38. Khan, S. A., Higdon, N. R. & Meisner, K. D. Coronary vasorelaxation by nitroglycerin: involvement of plasmalemmal calcium-activated K^+ channels and intracellular Ca^{++} stores. *J. Pharmacol. Exp. Ther.* **284**, 838–846 (1998).
39. Tybulewicz, V. L., Crawford, C. E., Jackson, P. K., Bronson, R. T. & Mulligan, R. C. Neonatal lethality and lymphopenia in mice with a homozygous disruption of the c-abl proto-oncogene. *Cell* **65**, 1153–1163 (1991).
40. Joyner, A. L. in *Gene Targeting: A Practical Approach* Vol. xviii, 234 (IRL at Oxford Univ. Press, Oxford, New York, 1993).
41. Sambrook, J., Maniatis, T. & Fritsch, E. F. *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Laboratory, Cold Spring Harbor, New York, 1989).
42. Hogan, B. in *Manipulating the Mouse Embryo: A Laboratory Manual* Vol. xvii, 497 (Cold Spring Harbor Laboratory Press, Plainview, New York, 1994).
43. Berkowitz, R. D. *et al.* CCR5- and CXCR4-utilizing strains of human immunodeficiency virus type 1 exhibit differential tropism and pathogenesis in vivo. *J. Virol.* **72**, 10108–10117 (1998).
44. Knot, H. J. & Nelson, M. T. Regulation of arterial diameter and wall $[\text{Ca}^{2+}]$ in cerebral arteries of rat by membrane potential and intravascular pressure. *J. Physiol. (Lond.)* **508**, 199–209 (1998).
45. Nelson, M. T. *et al.* Relaxation of arterial smooth muscle by calcium sparks. *Science* **270**, 633–637 (1995).
46. Horn, R. & Marty, A. Muscarinic activation of ionic currents measured by a new whole-cell recording method. *J. Gen. Physiol.* **92**, 145–159 (1988).
47. Simpson, L. O. A mouse model of spontaneous renal hypertension, blood pressure, heart weight, kidney weight and proteinuria relationships in NZB $\times$ OUW F1 hybrid female mice. *Pathology* **12**, 347–357 (1980).
48. Han, J. *et al.* Age-related changes in blood pressure in the senescence-accelerated mouse (SAM): aged SAMPI mice manifest hypertensive vascular disease. *Lab. Anim. Sci.* **48**, 256–263 (1998).

Acknowledgements

We thank B. K. Kobilka of Stanford University for advice; Y. Chen in the Stanford Transgenic Mouse facility for the blastocyst injections; G. Maw for advice and assistance on isolated artery images; S. Brett Welsh for genotyping and technical assistance; D. Hill-Eubanks for comments on the manuscript and D. P. Regula for paraffin sections of hearts. DHS-1 was provided by Merck Research Laboratories. This work was supported by a grant from the Federation for Anesthesia Education and Research to A.J.P. and by grants from the National Institutes of Health (H.L.B.I. and D.D.K.), National Science Foundation, Totman Medical Trust for Cerebrovascular Research to M.T.N., and an American Heart Association Fellowship to G.P. R.W.A. is an investigator with the Howard Hughes Medical Institute.

Correspondence and requests for materials should be addressed to R.W.A. (e-mail: raldrich@leland.stanford.edu).

Molecular emission from single-bubble sonoluminescence

Yuri T. Didenko, William B. McNamara III & Kenneth S. Suslick

Department of Chemistry, University of Illinois at Urbana-Champaign, 600 S. Mathews Avenue, Urbana, Illinois 61801, USA

Ultrasound can drive a single gas bubble in water into violent oscillation; as the bubble is compressed periodically, extremely short flashes of light (about 100 ps) are generated with clock-like regularity^{1–4}. This process, known as single-bubble sonoluminescence, gives rise to featureless continuum emission^{4,5} in water (from 200 to 800 nm, with increasing intensity into the ultra-violet). In contrast, the emission of light from clouds of cavitating bubbles at higher acoustic pressures (multi-bubble sonoluminescence¹) is dominated by atomic and molecular excited-state emission^{6–11} at much lower temperatures⁶. These observations have spurred intense effort to uncover the origin of sonoluminescence and to generalize the conditions necessary for its creation. Here we report a series of polar aprotic liquids that generate very strong single-bubble sonoluminescence, during which emission from molecular excited states is observed. Previously, single-bubble sonoluminescence from liquids other than water has proved extremely elusive^{12,13}. Our results give direct proof of the existence of chemical reactions and the formation of molecular excited states during single-bubble cavitation, and provide a spectroscopic link between single- and multi-bubble sonoluminescence.

The origin of the featureless emission of single-bubble sonoluminescence (SBSL) in water has become the subject of many theoretical speculations, and several mechanisms have been proposed, including black-body radiation⁵, bremsstrahlung^{14,15}, ion-electron recombination¹⁵, and confined electrons¹⁶. Regardless of the detailed mechanism, a consensus has emerged that SBSL involves extraordinary temperatures inside the bubble^{14–16}.

Whereas SBSL from water has been extensively studied, SBSL from other liquids in air has not¹². For a single cavitating bubble in water, Lohse and co-workers have argued convincingly that N₂ and O₂ are removed from the bubble's interior by reaction to form NO_x (which dissolves in and reacts with the bulk water), leaving primarily monatomic argon inside the bubble^{17,18}. In addition, sonolysis of water (to H atoms and OH radicals) is highly reversible. These properties make water a nearly perfect liquid for SBSL under air: cavitation-induced reactions do not yield gaseous polyatomic products that would hinder compressional heating of the bubble.

The only previous report¹³ of SBSL from liquids other than water found a weak emission from organic liquids (mostly alcohols), but only for unstable, moving xenon bubbles (*n*-butanol in air apparently also gives a very weak emission about 200 times less intense than water¹³). The spectra of these liquids under Xe again are featureless continua (we note, however, that emission from an excited state of C₂ was observed from a xenon bubble pulsating asymmetrically on a thermocouple wire in dodecane¹⁹). In contrast to water, sonolysis of other liquids produces small polyatomic gases⁸ during the entropy-driven, high-energy chemistry that occurs within the bubble. We believe that build-up of such gases inside a single cavitating bubble precludes the heating necessary to generate SBSL.

From this hypothesis, we predict that a liquid will support SBSL under air if it has (1) a low vapour pressure (necessary to minimize the amount of gas formed during sonolysis) and (2) a high content of heteroatoms such as O or N (to form species that dissolve in or react with the surrounding liquid, and thus are removed from the

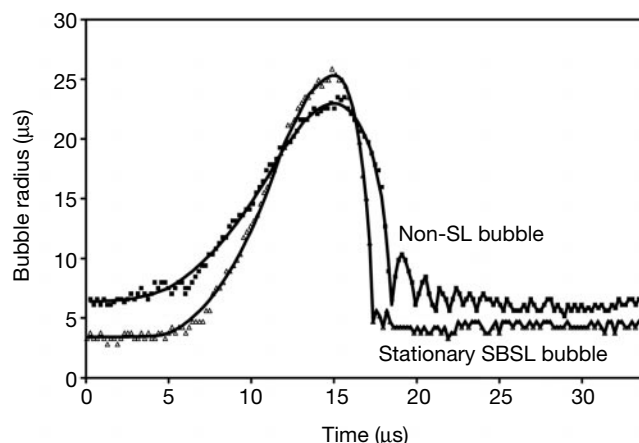


Figure 1 Radius versus time curves for single-bubble cavitation in methylformamide. Data are shown slightly below and slightly above the sonoluminescence (SL) threshold (acoustic pressure, ~1.1 bar). The radius data (shown as data points) were generated from stroboscopic imaging, so the time axis is relative and does not represent the phase of the bubble with respect to the acoustic field. The SBSL apparatus consisted of a spherical, quartz, 100-ml cell with a cylindrical piezoceramic transducer cemented to the bottom. The transducer was driven by a function generator and amplifier at a resonant frequency of about 30 kHz. A calibrated needle hydrophone (DAPCO) was used to measure the approximate pressure amplitude at the bubble location. The stroboscopic imaging system was similar to that used in ref. 21. The image of the bubble was recorded through a microscope by a Pulnix 9701 CCD camera. The bubble was illuminated by a light-emitting diode, which was driven by a pulse generator (HP 8116A). One of the outputs from HP 8904A signal synthesizer was used to trigger the pulse generator and the other sent to the power amplifier that drove the SBSL cell.

bubble). The kinetics of the reactions that form emissive excited states (of, for example, C₂ or CN) may also be important, given the short period of emission during SBSL. Using these criteria, we have discovered intense SBSL in air from numerous polar, aprotic liquids, including formamide (H₂NCHO), *N*-methylformamide ((H₃C)HNCHO), *N,N*-dimethylformamide ((H₃C)₂NCHO), *N*-methylacetamide ((H₃C)HNC(CH₃)O), 1,2-diaminoethane (H₂NCH₂CH₂NH₂), dimethylsulphoxide ((H₃C)₂SO) and adiponitrile (NC(CH₂)₄CN). In keeping with our prediction, the intensity of sonoluminescence in these liquids can be comparable to, or even higher than, that of water. We note that both constraints—that is, (1) and (2) above—on the liquid properties must be met: SBSL is not observed from silicone oil or long-chain aliphatic alcohols (which have very low volatility, but low heteroatom content) in air, nor from formic acid (which has a high heteroatom content, but high volatility).

Both the dynamics of bubble motion in these organic liquids and their SBSL spectra differ significantly from that seen in water. We will discuss the bubble behaviour of *N*-methylformamide in detail. A bubble can be trapped at low acoustic pressure ($P_A \approx 0.9$ bar for *N*-methylformamide) in the centre of the flask (see Fig. 1 legend for details), where it is spatially stationary and pulsates without emitting light. As P_A is increased, the bubble starts to move around the flask's centre and sonoluminescence is first observed (at $P_A \approx 1.3$ bar for *N*-methylformamide). The velocity of the bubble motion at this point is about 3 mm s⁻¹, and its trajectory is often circular or elliptical. The light intensity and the bubble movement both increase with increasing P_A . Microbubble shedding²⁰ or bubble-induced detuning may be responsible for such bubble motion, but additional work is required to explain this phenomenon.

We will refer to light emission from a non-stationary bubble as moving single-bubble sonoluminescence, M-SBSL. These M-SBSL bubbles can be maintained at maximum emission intensities for several minutes. The maximum emission intensities (relative to water at 1.0) are adiponitrile, 2.0; formamide, 1.2; methylforma-

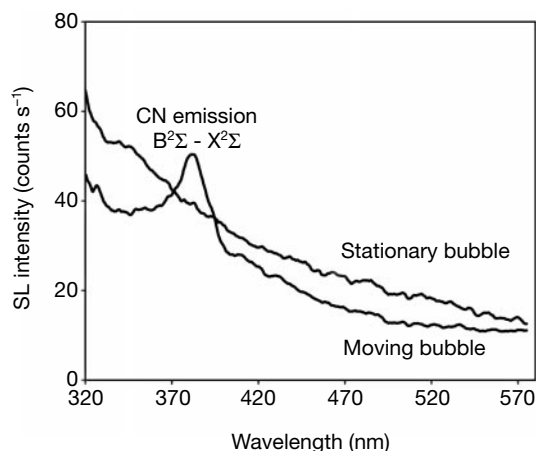


Figure 2 SBSL spectra of methylformamide at an acoustic pressure of ~ 1.1 bar for both a stationary and a moving single bubble. The excited-state CN comes from the liquid vapour rather than from the nitrogen gas initially present inside the bubble, because (1) M-SBSL spectra under a partial pressure of Ar (in the absence of N_2 or O_2) show greater CN emission; and (2) M-SBSL spectra from dimethylsulphoxide (which contains no nitrogen) in air do not exhibit CN emission. SBSL spectra of methylformamide were collected at 295 K under 10 torr of air, using a Jobin-Yvon Triax 320 monochromator with a Spectrum One CCD detector (Instrument SA, $1,024 \times 256$ array). Spectra were corrected for both the light absorption by the solutions and the response of the optical system against standard lamps (which could be traced back to NIST standards).

mide, 0.14; dimethylsulphoxide, 0.14; dimethylformamide, < 0.05 ; and 1,2-diaminoethane, < 0.05 ; intensities were measured in the same cell with the spectrograph in zeroth order. These liquids all absorb light much more strongly in the ultraviolet region than water does, so the relative intensities of emission can only be taken as an approximate measure of the efficiency of SBSL. Above about 1.5–1.6 bar in *N*-methylformamide, the bubble cannot be held at the centre of the flask and disappears; such behaviour is typical for all liquids reported here.

Hysteresis is observed for SBSL in these liquids. By decreasing the amplitude of ultrasound reaching a moving emitting bubble, it was possible to get light emission at lower P_A than the initial threshold of sonoluminescence inception for all of our non-aqueous liquids. In the cases of formamide and methylformamide, upon decreasing the acoustic amplitude to about 1.1 bar, the bubble becomes stationary and still emits; we will refer to this as stationary single-bubble sonoluminescence, S-SBSL. The intensity of sonoluminescence under the conditions of S-SBSL was approximately the same as, or a little higher than, the emission intensity during M-SBSL at the same acoustic pressure. S-SBSL can be obtained only by first establishing a bubble showing M-SBSL and then reducing P_A .

Stroboscopic imaging²¹ revealed that the stationary emitting bubble pulsates regularly with only a few, weak after-bounces. In contrast, a non-emitting stationary bubble is much larger in size and shows strong, multiple afterbounces (Fig. 1). This difference in dynamics between sonoluminescent and non-sonoluminescent bubbles has also been observed in water, and is consistent with gas removal from the bubble by chemical reactions during SBSL^{18,21}.

The emission spectra from these liquids are especially revealing. In contrast to the featureless spectrum of water SBSL, M-SBSL from adiponitrile, formamide and methylformamide all show relatively strong emission from excited-state CN ($B^2\Sigma^+ - X^2\Sigma^+$). At or near the minimum P_A necessary for M-SBSL, this molecular emission is superimposed on an underlying continuum (Figs 2, 3). As the P_A increases, the continuum increases in intensity until the CN emission is no longer discernible (Figs 3, 4). The bubble hysteresis made it possible to collect spectra from *N*-methylformamide for both S-SBSL (which is featureless) and M-SBSL (which

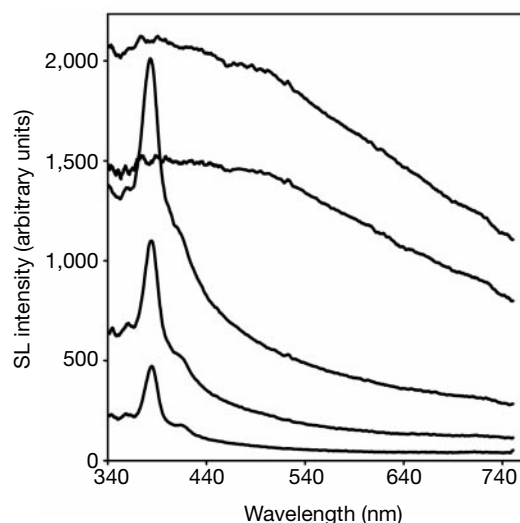


Figure 3 M-SBSL spectra of adiponitrile. Acoustic pressure increases from bottom to top, from 1.7 to 1.9 bar.

shows CN emission) under essentially identical acoustic conditions (Fig. 2).

We propose that the difference between S-SBSL and M-SBSL spectra at similar acoustic pressure lies in the sphericity and severity of the bubble collapse and the effect this has on the bubble contents. A moving bubble cannot collapse as symmetrically as a stationary bubble; it is therefore likely that the effective temperatures reached during S-SBSL are much higher than during M-SBSL. The observation of emission from CN excited states during M-SBSL shows that the temperature is below about 15,000 K under these conditions, for at least part of the emitting bubble; above this temperature the CN (whose dissociation energy is 7.8 eV) would not be stable long enough to emit²². In the case of S-SBSL (which is presumably at higher temperatures than M-SBSL), CN is either thermally dissociated or no longer is formed within the collapsing bubble, and CN emission is not observed.

Several experimental parameters are found to affect the intensity and stability of SBSL in these liquids. Vapour pressure (which plays an essential role in the intensity and temperature of MBSL⁶) appears to be important for SBSL: all of our successful non-aqueous liquids have very low volatilities. For example, adiponitrile (vapour pressure < 0.1 torr at 22 °C; ref. 23) gives intense SBSL, whereas acetonitrile (CH_3CN , vapour pressure 90 torr), a liquid very similar in chemical properties to adiponitrile, fails to give any SBSL. Similarly, the intensity of SBSL decreases more than 20 times from formamide (vapour pressure 0.05 torr at 22 °C) to methylformamide (0.2 torr at 22 °C), to dimethylformamide (3.2 torr at 22 °C). M-SBSL occurs in *N*-methylformamide equilibrated with air pressures from 10 to 150 torr; as the gas pressure increases, M-SBSL occurs at higher acoustic pressures and the bubble's spatial stability decreases.

The acoustic pressure during M-SBSL significantly affects the emission spectra (Fig. 3). As P_A is increased during M-SBSL, the intensity of bubble collapse will also increase, leading to higher temperatures and an initial increase in CN emission as more CN excited states are formed. As the P_A is further increased, however, CN is either thermally dissociated or no longer formed within the collapsing bubble, and a sharp decrease in the intensity of CN emission is observed. Subsequent experiments have shown that the transition between the absence and presence of CN emission occurs on a timescale of about 0.1 s. This is further evidence that CN emission comes from sonolysis of vapour of the liquid, rather than dissolved N_2 , because N_2 diffusion into the bubble requires several

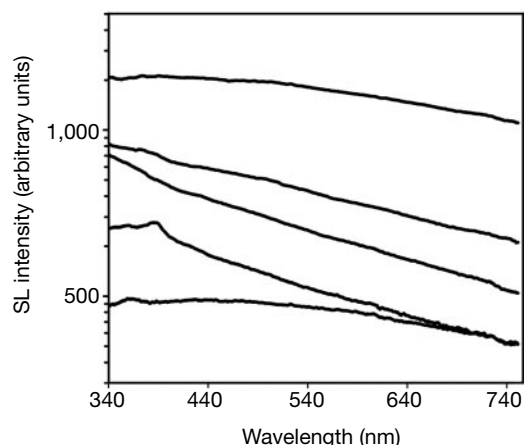


Figure 4 M-SBSL spectra of adiponitrile, formamide, water, methylformamide and dimethylsulphoxide (from the top to the bottom of the figure). The acoustic pressure was 1.4 bar for all liquids except adiponitrile, where it was ~ 1.9 bar. High-purity non-aqueous liquids ($> 99\%$ purity) were used as purchased. Water was purified by ion exchange (to $18\text{ M}\Omega\text{ cm}$) and filtration through $0.2\text{ }\mu\text{m}$ pore filters.

seconds to replenish¹⁷. The observation of CN emission during SBSL also proves that chemical reactions are taking place during SBSL, as has been suggested from hydrodynamic modelling of single-bubble cavitation^{17,18}.

The M-SBSL spectra of dimethylsulphoxide and adiponitrile at high acoustic intensities are significantly different from those of other liquids (Fig. 4): their intensity is essentially flat above 380 nm and cannot be fitted to black-body emission at a single temperature. Because there is no nitrogen in dimethylsulphoxide, there is no CN emission observed. One might have hoped to see band emission from S_2 , SO , SO_2 or other species that should be produced from sonolysis of dimethylsulphoxide. None of these species, however, is as stable as CN at high temperatures. For reference, the dissociation energies for SO and CN are 5.4 and 7.8 eV, respectively²³.

The origin of the continuum in SBSL remains an open question. There are two likely contributors: (1) the sum of broadened emission from multiple molecular and atomic sources, and (2) bremsstrahlung emission from partial ionization of the heated gas within the bubble. The continuum in flames is generally assigned to the former mechanism²⁴. For example, broadened CO emission could be a significant ultraviolet component of SBSL spectra as it has a dissociation energy considerably higher than that of CN (11.1 eV; refs 24, 25). Emission from CO_2 , N_2O , NO and NO_2 (refs 24–26) may also be contributors to the continuum. Recent experiments on the timing of the SBSL flash from water²⁷ showed that the emission at long wavelengths can last slightly longer than that at short wavelengths; this difference could originate from the varying emissivities and lifetimes of multiple emitting species as a function of wavelength and internal temperature.

The other likely component of the continuum emission is bremsstrahlung radiation from electron–neutral and electron–ion interactions due to partial ionization of the heated gas within the bubble. Calculations show that lower-temperature bremsstrahlung (approximately 12,000 K) could give wavelength-independent intensities similar to those we observe in dimethylsulphoxide (Fig. 4; P.L. Hagelstein, personal communication). Under conditions of maximum collapse efficiency (that is, stationary single-bubble cavitation, minimum solvent vapour pressure, Xe gas within the bubble, and so on), the temperature within the cavity increases, and bremsstrahlung emission is expected to increase. From comparison of our experimental data to Hagelstein's calculations, the

wavelength dependence of S-SBSL from *N*-methylformamide is consistent with an effective bremsstrahlung emission temperature of approximately 30,000 K.

We believe that M-SBSL represents less extreme conditions ($< 15,000\text{ K}$) than those present during S-SBSL, due to the less-effective collapse of moving bubbles. M-SBSL therefore provides a long-sought spectroscopic link between the phenomena of single-bubble cavitation (which occurs at low acoustic pressures) and bubble-cloud cavitation (which occurs at high acoustic pressures). As the sphericity of bubble collapse increases (from MBSL to M-SBSL to S-SBSL), the efficacy of compression increases, the effective temperatures within the bubble increase, and the emission changes from spectra dominated by excited-state molecular emission to featureless (bremsstrahlung-like) spectra. □

Received 2 June; accepted 30 August 2000.

- Suslick, K. S. & Crum, L. A. in *Encyclopedia of Acoustics* (ed. Crocker, M. J.) 271–282 (Wiley-Interscience, New York, 1997).
- Gaitan, D. F., Crum, L. A., Church, C. C. & Roy, R. A. Sonoluminescence and bubble dynamics for a single, stable, cavitation bubble. *J. Acoust. Soc. Am.* **91**, 3166–3183 (1992).
- Gompf, B. *et al.* Resolving sonoluminescence pulse width with time-correlated single photon counting. *Phys. Rev. Lett.* **79**, 1405–1408 (1997).
- Hiller, R. A., Putterman, S. J. & Weninger, K. R. Time-resolved spectra of sonoluminescence. *Phys. Rev. Lett.* **80**, 1090–1093 (1998).
- Hiller, R., Putterman, S. J. & Barber, B. P. Spectrum of synchronous picosecond sonoluminescence. *Phys. Rev. Lett.* **69**, 1182–1184 (1992).
- McNamara, W. B. III, Didenko, Y. T. & Suslick, K. S. Sonoluminescence temperatures during multi-bubble cavitation. *Nature* **401**, 772–775 (1999).
- Suslick, K. S. & Flint, E. B. Sonoluminescence from non-aqueous liquids. *Nature* **330**, 553–555 (1987).
- Flint, E. B. & Suslick, K. S. Sonoluminescence from nonaqueous liquids: emission from small molecules. *J. Am. Chem. Soc.* **111**, 6987–6992 (1989).
- Flint, E. B. & Suslick, K. S. Sonoluminescence from alkali metal salt solutions. *J. Phys. Chem.* **95**, 1484–1488 (1991).
- Suslick, K. S., Flint, E. B., Grinstaff, M. W. & Kemper, K. Sonoluminescence from metal carbonyls. *J. Phys. Chem.* **97**, 3098–3099 (1993).
- Didenko, Y. T. & Pugach, S. P. Spectra of water sonoluminescence. *J. Phys. Chem.* **98**, 9742–9749 (1994).
- Gaitan, D. F. *et al.* Spectra of single-bubble sonoluminescence in water and glycerine-water mixtures. *Phys. Rev. E* **54**, 525–528 (1996).
- Weninger, K. *et al.* Sonoluminescence from single bubbles in nonaqueous liquids: new parameter space for sonochemistry. *J. Phys. Chem.* **99**, 14195–14197 (1995).
- Moss, W. C., Clarke, D. B. & Young, D. A. Calculated pulse widths and spectra of a single sonoluminescing bubble. *Science* **276**, 1398–1401 (1997).
- Hilgenfeldt, S., Grossmann, S. & Lohse, D. A simple explanation of light emission in sonoluminescence. *Nature* **398**, 402–405 (1999).
- Bernstein, L. S. & Zakin, M. R. Confined electron model for single-bubble sonoluminescence. *J. Phys. Chem.* **99**, 14619–14627 (1995).
- Lohse, D. *et al.* Sonoluminescing air bubbles rectify argon. *Phys. Rev. Lett.* **78**, 1359–1362 (1997).
- Lohse, D. & Hilgenfeldt, S. Inert gas accumulation in sonoluminescing bubbles. *J. Chem. Phys.* **107**, 6986–6997 (1997).
- Weninger, K. R. *et al.* Sonoluminescence from an isolated bubble on a solid surface. *Phys. Rev. E* **56**, 6745–6749 (1997).
- Matula, T. J. & Crum, L. A. Evidence for gas exchange in single-bubble sonoluminescence. *Phys. Rev. Lett.* **80**, 865–868 (1998).
- Ketterling, J. A. & Apfel, R. E. Experimental validation of the dissociation hypothesis for single bubble sonoluminescence. *Phys. Rev. Lett.* **81**, 4991–4994 (1998).
- Bernstein, L. S., Zakin, M. S., Flint, E. B. & Suslick, K. S. Cavitation thermometry using molecular and continuum sonoluminescence. *J. Phys. Chem.* **100**, 6612–6619 (1996).
- Lide, D. R. (ed.) *CRC Handbook of Chemistry and Physics* (CRC Press, Boca Raton, 1993–1994).
- Gaydon, A. G. *The Spectroscopy of Flames* 2nd edn (Chapman and Hall, London, 1974).
- Hertzberg, G. *Molecular Spectra and Molecular Structure: Electronic Spectra and Electronic Structure of Polyatomic Molecules* (Van Nostrand, New York, 1966).
- Zel'dovich, Y. B. & Raizer, Yu. P. *Physics of Shock Waves and High Temperature Hydrodynamic Phenomena* (Academic, New York, 1966).
- Moran, M. J. & Sweider, D. Measurements of sonoluminescence temporal pulse shape. *Phys. Rev. Lett.* **80**, 4987–4990 (1998).

Acknowledgements

We thank T. J. Matula for the loan of a calibrated needle hydrophone and many discussions, and S. Kulikov for assistance in constructing the imaging system. This work was supported by the US Defense Advanced Research Projects Agency and in part by the National Science Foundation.

Correspondence and requests for materials should be addressed to K.S.S. (e-mail: ksuslick@uiuc.edu).

Strain effects and phase transitions in photonic resonator crystals

Harald Pier^{*‡}, Eli Kapon^{*} & Michael Moser[†]

^{*} Department of Physics, Swiss Federal Institute of Technology (EPFL), 1015 Lausanne, Switzerland

[†] Avalon Photonics, Badenerstrasse 569, 8048 Zurich, Switzerland

Optical structures with periodic variations of the dielectric constant in one or more directions (photonic crystals¹) have been employed extensively for studying optical diffraction phenomena. Practical interest in such structures arises from the possibilities^{2–4} they offer for tailoring photon modes, and thereby the characteristics of light propagation and light–matter interactions. Photonic resonator crystals comprising two-dimensional arrays of coupled optical microcavities have been fabricated using vertical-cavity surface-emitting laser wafers⁵. In such structures, the light propagates mostly normal to the periodic plane. Therefore, the corresponding lateral Bragg-periodicities are larger, a feature that is advantageous for device manufacture as it allows for larger lattice constants in the lateral direction. Here we investigate strain effects in a photonic resonator crystal by shifting neighbouring lattice rows of microcavities in opposite directions, thereby introducing an alternating square or quasi-hexagonal pattern of shear strain. We find that, for strain values below a critical threshold, the lasing photon mode is virtually locked to the corresponding mode supported by the unstrained photonic crystal. At the critical strain value, we observe a phase-transition-like switching between the square and quasi-hexagonal lattice modes. The tolerance of subcritical strains suggests that the resonator crystal may be useful for applications requiring high spatial coherence across the lattice, while the mode switching could potentially be exploited in free-space optical communications.

Periodic variations in the refractive index of a dielectric medium can yield photonic crystals whose optical properties show analogies with the electronic properties of atomic crystals¹. The formation of a photonic band structure in such photonic crystals relies on coherent, Bragg reflection of the electromagnetic modes off the periodic lattice. By carefully choosing the lattice type and the refractive-index contrast, a full photonic bandgap—that is, an optical frequency range in which no propagating photon modes exist for any direction of propagation—can be obtained^{2–4}. Photonic crystal structures exhibiting such a full (three-dimensional, 3D) photonic bandgap are the optical equivalents of semiconductor material with its electronic bandgap. Their optical properties can be tailored on the scale of the optical wavelength in the material; for example, by adding optical impurities (that is, replacing a crystal constituent by an element with a different refractive index). Two-dimensional (2D) photonic bandgap material has been used to realize a waveguiding scheme in planar optical waveguides with channels defined by periodic hole structures^{6,7}. In this case, the electromagnetic mode mainly propagates in the plane of periodicity. Periodic arrays of voids along an optical fibre can modify dramatically the modal structure of that fibre⁸. Here, the electromagnetic mode propagates mostly normal to the periodic plane with a main component $k_{\parallel}$ of the wavevector that does not undergo any Bragg reflection. The fact that only a (small) component of $\mathbf{k}$ undergoes Bragg reflections may lead to photonic crystal designs with periodicities longer than the wavelength in the material.

Closely related to this last category are periodic arrays of coupled optical waveguides or resonators (Fig. 1, left). In this case, the waveguides or resonators are arranged in a periodic array in a plane normal to the main propagation direction. The periodicity matches the Bragg condition for the corresponding transverse component $k_{\perp}$ of the wavevector. For a first-order Bragg reflection this translates into a lateral periodicity $\Lambda = \lambda/2\sin\theta$, where θ is the angle between the propagation direction and the guiding axis. In the case of such resonator crystals, the resulting 2D photon modes exhibit intensity peaks at each resonator; within a given photon mode, the electromagnetic field is coherent across the entire lattice. Examples of such coupled resonator structures include ‘photonic molecules’ made using etched Bragg optical microcavities⁹ and coupled arrays of vertical-cavity surface-emitting lasers (VCSELs)⁵.

Here we report the observation of strain effects and phase transition behaviour in photonic crystals comprising 2D arrays of coupled laser resonators. The highly uniform semiconductor VCSEL wafers utilized were grown using organometallic chemical vapour deposition, and were designed to lase at 940-nm wavelength at room temperature (Fig. 1, right). Details of the wafer structure are given in ref. 10. The extremely low round-trip gain in these lasers requires mirror reflectivities close to 100% for attaining the lasing threshold, and therefore slight (of the order of 1%) reflectivity variations suffice to define the position of the lasing cavity on the VCSEL wafer. Such variations are achieved by depositing a gold film on a Cr grid on top of the upper VCSEL mirror, and therefore yield 2D crystals of high-finesse resonators defined by the gold pixels⁵. The close proximity of these Au pixels (with as small as 1- μm spacing) promotes optical coupling between adjacent microcavities via diffraction of the optical field at the edges of the pixels. This coupling results in phase-locking of the modes of individual cavities into collective modes (‘supermodes’) of the entire photonic resonator crystal. The lithographic definition of the lattices allows the realization of numerous one-dimensional (1D) or 2D lattice arrangements, limited only by the resolution of the lithographic method employed. Such optical structures have also been used recently in studies of optical disorder in 1D or 2D photonic crystals^{10,11}.

Strain can be introduced into such photonic crystals by properly designing the Au pixel arrangement, in analogy with strained atomic crystals for which deviations in the atomic positions from that of a regular lattice are present. The deviation of the atomic positions from the regular sites of a given lattice configuration is then described by a displacement field $\mathbf{u}(x_i)$ which, in turn, defines the strain tensor $S_{ij} = 1/2[(\partial u_i/\partial x_j) + (\partial u_j/\partial x_i)]$, where the x_i ($i = 1, 2, 3$) denote the three axes in a cartesian coordinate system. In the case of an atomic crystal, the effect of the strain field on the band structure of the electronic excitations is formulated in terms of the corresponding deformation potential of the lattice¹². In photonic crystals supporting lasing photon modes, such as those discussed here, the strain is expected to modify the cavity optical gain associated with each collective mode and will thereby determine the supermode that will eventually lase.

We have fabricated a series of ‘alternating-shear strain’, square-lattice VCSEL arrays consisting of square pixels (4- μm pixel width and $a = 5 \mu\text{m}$ lattice constant), in which every other row is shifted by $\delta = 0.5 \mu\text{m}$, $1.0 \mu\text{m}$, $1.5 \mu\text{m}$, $2.0 \mu\text{m}$ or $2.5 \mu\text{m}$. A square lattice structure with no pixel shifts was also made on the same wafer for control purposes. The amplified spontaneous emission (ASE) patterns of these structures, measured below lasing threshold through the bottom mirror, exhibit brighter emission at each pixel. This is because the cavity finesse there is higher than it is at the Cr-coated parts, which allows a direct visualization of the photonic crystal layout (see Fig. 2, top row). The strain tensor for this particular structure is given by $S_{11} = S_{22} = 0$, $S_{12} = S_{21} = \pm\delta/2a$ (+/– sign for odd/even rows).

[‡] Present address: Swiss Federal Institute of Intellectual Property, Einsteinstrasse 2, CH-3003, Bern, Switzerland.

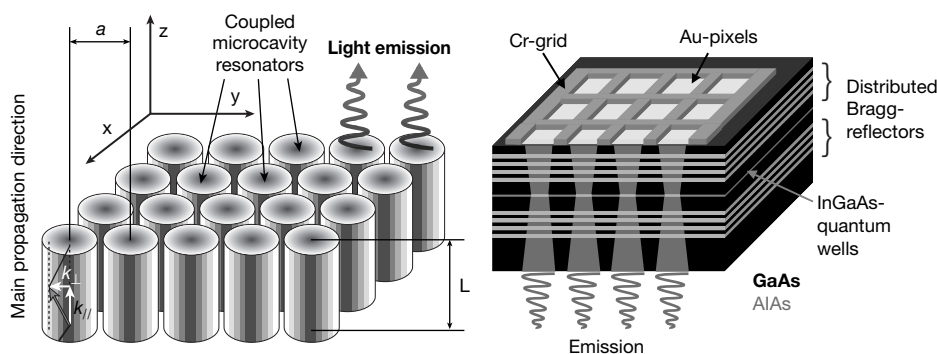


Figure 1 Photonic resonator crystals realized using coupled arrays of vertical-cavity surface-emitting lasers, VCSELs. Left, schematic illustration of a photonic resonator

crystal. Right, schematic illustration of a 2D photonic resonator crystal realized using coupled arrays of VCSELs. Lasing is observed from the bottom mirror.

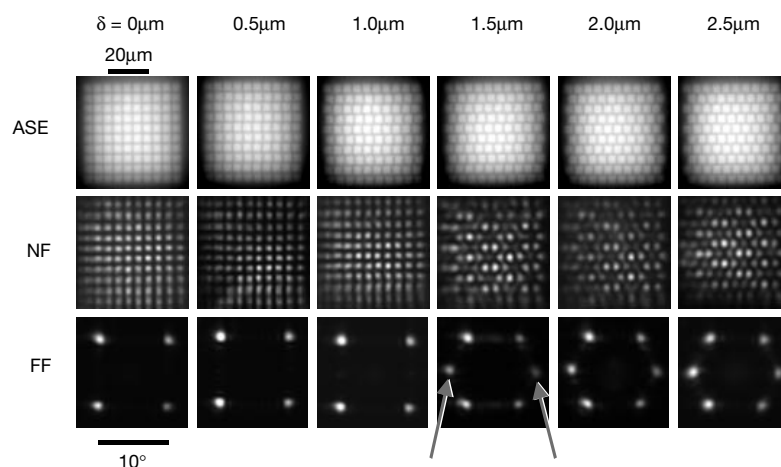


Figure 2 Optical-field patterns of alternating-shear strained photonic resonator crystals. The figure shows measured amplified spontaneous emission (ASE), and near-field (NF) and far-field (FF) patterns for different displacement parameters δ .

Above threshold, the near-field (NF) intensity patterns (Fig. 2, middle row) show intense emission from each lattice site, with intensity nulls in between adjacent pixels. The corresponding far-field (FF) patterns (Fig. 2, bottom row), essentially the Fourier transform of the near-field distribution, contain information regarding the phase relationships among the optical fields at the different lattice sites. Comparison with a coupled-mode model of the structures¹³ shows that the lasing just above threshold is exclusively due to the excitation of the highest order array supermode (located at the edge of the photonic band), for which nearest-neighbour lattice sites have π radians phase difference. This phase difference, resulting in vanishing intensity between the lattice sites, minimizes the cavity loss of that particular supermode. The characteristic phase distribution of the highest-order supermode yields the diffraction-limited, four-lobed beam shape shown in the left-most column of Fig. 2. Thus, above lasing threshold the efficient optical coupling gives rise to a highly (spatially and temporally) coherent optical field distributed across the entire 2D photonic crystal. We note that in the case of the hexagonal-lattice structure, it is impossible to obtain simultaneously a π phase difference between all nearest-neighbour pairs; this 'phase frustration' results in illumination of only part of the lattice. This explains the observed honeycomb near-field patterns for the case $\delta = 2.5 \mu\text{m}$, which represents a quasi-hexagonal lattice (rightmost column in Fig. 2).

For lattice displacements $\delta \leq 1 \mu\text{m}$, the observed mode-peak displacements are very small, and the supermode virtually 'locks' to the unstrained lattice positions. For $\delta \geq 1.5 \mu\text{m}$, the optical field

abruptly switches to a different supermode, related to the quasi-hexagonal lattice obtained for $\delta = 2.5 \mu\text{m}$. This new supermode stays locked to the quasi-hexagonal lattice mode for the entire $\delta \geq 1.5 \mu\text{m}$ range, similar to the locking to the square lattice for $\delta \leq 1 \mu\text{m}$. Experiments with finer displacements show that the switching occurs within less than $0.1 \mu\text{m}$ variations in δ near $\delta = 1 \mu\text{m}$. The locking to the square lattice is better visualized in Fig. 3, which shows the near-field lasing pattern superimposed on the amplified spontaneous emission pattern for the lattice with $\delta = 0.9 \mu\text{m}$. The abrupt switching of the supermode pattern is summarized quantitatively in Fig. 4, which depicts the observed displacement of the peak intensities of each pixel as a function of the lattice displacement δ (displacements were measured near the centre of the lattice). The evolution of the near-field patterns as a function of the displacement is accompanied by a corresponding evolution of the far-field pattern (bottom row in Fig. 2). This evolution is characterized by the shrinking of the distance between the lobes in the square lattice, in the direction parallel to the displacement, and the appearance of two additional lobes, marked by the two arrows in Fig. 2. In fact, these two additional lobes are present already in the structures with $\delta \leq 1 \mu\text{m}$; however, their intensity is too dim to be observed in Fig. 2. Moreover, the intensity of these additional lobes can be used as a measure of the actual displacement of the peaks in the near-field pattern. This was accomplished by using coupled-mode theory to calculate the eigenmodes of the strained structures as a function of lattice displacement. Fitting of the intensities of the additional lobes in the calculated far-field

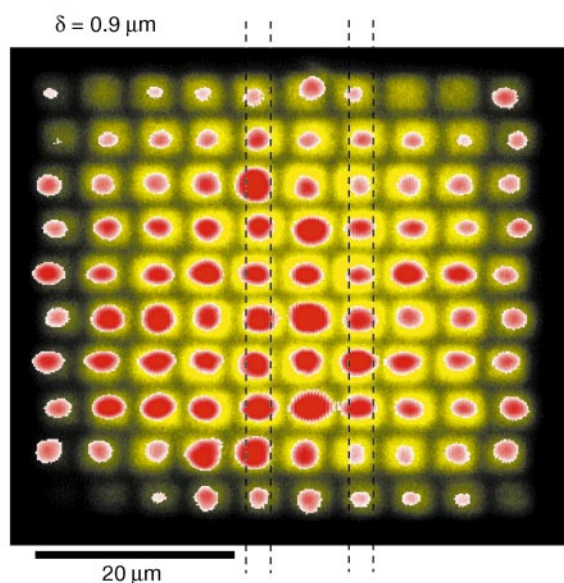


Figure 3 Near-field pattern of a square-lattice photonic crystal incorporating subcritical strain. The amplified spontaneous emission pattern (in yellow) shows the positions of the lattice sites, and is superimposed on the near-field pattern measured above threshold (in red). The image illustrates the locking of the near-field pattern to the unstrained pattern for an alternating-shear strained square lattice with a displacement parameter $\delta = 0.9 \mu\text{m}$.

patterns to the experimental values by adjusting the coupling coefficients and the peak displacements allows the determination of the expected peak displacements in the near-field distribution. These calculated peak displacements, obtained from the measured far-field intensity distributions, are also shown in Fig. 4; good agreement is obtained with the displacements measured directly from the near fields.

The observed effects of strain on the optical field in these photonic crystals can be understood by considering the mode selection mechanisms in a lasing photonic crystal. The multitude of modes in a given lattice structure, forming the photonic band, can be ordered according to the imaginary part of their propagation constant, or their optical cavity losses. The peculiar feature of a lasing photonic crystal is that only the lowest loss modes are sustained above lasing threshold. In fact, all the measurements reported here have been performed near threshold, such that a single photon mode (the one with the lowest optical cavity loss) oscillates. In other words, the lasing mode in these photonic crystals is always located at the photonic band edge. In the strain-free, square lattice structure, the out-of-phase mode oscillates since the π phase difference between neighbouring pixels ensures intensity nulls at the low-reflectivity Cr grid, thus minimizing its modal losses. As alternating shear strain is introduced, the mode pattern virtually locks to the strain-free square lattice out-of-phase mode that remains the lowest cavity mode of the system¹⁵. However, the entire band structure of the photonic crystal evolves with increasing strain. In particular, the cavity loss of the lasing mode increases due to the reduced overlap between the mode intensity peaks and the high-reflectivity Au pixels. For sufficiently large displacements, the modal loss of the square lattice out-of-phase mode surpasses that of the quasi-hexagonal mode and the latter (still at the edge of the modified photonic band) takes over. This observed behaviour of the photon modes is reproduced by a coupled-cavity model¹⁴ that treats these photonic crystals as coupled resonators¹⁵.

At the critical displacement value where the abrupt mode switching occurs, the observed square-lattice and the quasi-hexagonal lattice supermodes have the same modal loss¹⁵. For lower (higher) values, the square- (quasi-hexagonal-) lattice mode exhibits the

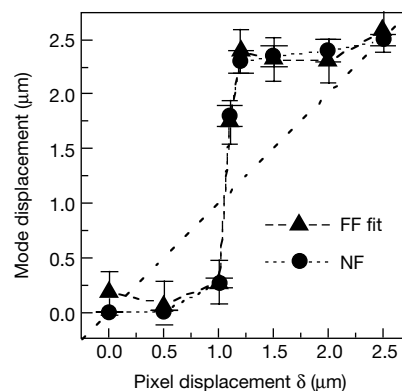


Figure 4 Displacement of optical field as function of lattice displacement. The figure shows measured displacements in the optical mode peak positions as a function of the displacement of the lattice sites (circles). The fitted mode displacement, inferred from the intensity of the additional lobes appearing in the far-field patterns, is also shown (triangles). The dashed line represents a mode displacement that would completely follow the lattice displacement.

lowest modal loss and is therefore the lasing mode. Thus, as the lattice strain is tuned across its critical displacement value, the system crosses from one stable collective mode to another. We stress the fact that there is no gradual evolution of a 'square-type' collective mode towards a 'quasi-hexagonal type', but rather an abrupt switching. This behaviour is reminiscent of a first-order phase transition in a thermodynamic system.

We have shown that these photonic resonator crystals can tolerate subcritical amounts of strain, which results in stable, coherent lasing modes even in the presence of structural disorder. This makes such laser array structures useful for applications requiring a high degree of spatial coherence with specific amplitude and phase distributions across the lattice. On the other hand, the abrupt switching of the lasing photon mode patterns could be exploited to achieve controlled beam switching and reconfiguration, which would be useful, for example, in free-space optical communication applications. Such dynamic beam switching could be achieved using injection locking of strained VCSEL arrays with δ near the critical displacement, as has recently been proposed and analysed¹⁶. □

Received 4 August 1999; accepted 23 August 2000.

- Joannopoulos, J. D., Villeneuve, P. R. & Fan, S. Photonic crystals: putting a new twist on light. *Nature* **386**, 143–149 (1997).
- Yablonovitch, E. Inhibited spontaneous emission in solid-state physics and electronics. *Phys. Rev. Lett.* **58**, 2059–2062 (1987).
- John, S. Strong localization of photons in certain disordered dielectric superlattices. *Phys. Rev. Lett.* **58**, 2486–2489 (1987).
- John, S. Localization of light. *Phys. Today* **32**, 40 (May 1991).
- Orenstein, M. et al. Two-dimensional phase-locked arrays of vertical-cavity semiconductor lasers by mirror reflectivity modulation. *Appl. Phys. Lett.* **58**, 804–806 (1991).
- Mekis, A. High transmission through sharp bends in photonic crystal waveguides. *Phys. Rev. Lett.* **77**, 3787–3790 (1996).
- Lin, S. Y. et al. Experimental demonstration of guiding and bending of electromagnetic waves in a photonic crystal. *Science* **282**, 274–276 (1998).
- Knight, J. C. et al. Photonic band gap guidance in optical fibers. *Science* **282**, 1476–1478 (1998).
- Bayer, M. et al. Optical modes in photonic molecules. *Phys. Rev. Lett.* **81**, 2582–2585 (1998).
- Golshani, A. et al. Photon mode localization in disordered arrays of vertical cavity surface emitting lasers. *J. Appl. Phys.* **85**, 2454–2456 (1999).
- Pier, H. & Kapon, E. Photon localization in lattices of coupled vertical-cavity surface-emitting lasers with dimensionalities between one and two. *Opt. Lett.* **22**, 546–548 (1997).
- Bir, G. L. & Pikus, G. E. *Symmetry and Strain-Induced Effects in Semiconductors* (Wiley, New York, 1974).
- Orenstein, M. et al. Large two-dimensional arrays of phase-locked vertical cavity surface emitting lasers. *Appl. Phys. Lett.* **60**, 1535–1537 (1992).
- Hardy, A. & Kapon, E. Coupled-mode formulations for parallel-laser resonators with application to vertical-cavity semiconductor-laser arrays. *IEEE J. Quant. Electron.* **32**, 966–971 (1996).
- Fishman, T. et al. Modal expansion analysis of strained photonic lattices based on vertical cavity surface emitting laser arrays. *Appl. Phys. Lett.* **74**, 3595–3597 (1999).

16. Fishman, T., Hardy, A. & Kapon, E. Mode switching in shear-strained and optically disordered photonic lattices of VCSEL by means of injection locking. *Appl. Phys. Lett.* **76**, 816–818 (2000).

Acknowledgements

This work was supported by the Fonds National Suisse de la Recherche Scientifique.

Correspondence and requests for materials should be addressed to H.P.
(e-mail: harald.pier@ipi.ch).

Origin of ferromagnetic exchange interactions in a fullerene–organic compound

Bakhyt Narymbetov*, Ales Omerzu†, Viktor V. Kabanov‡, Madoka Tokumoto‡, Hayao Kobayashi* & Dragan Mihailovic‡

* Institute for Molecular Science, Okazaki 444-8585, Japan

† Institute Josef Stefan, Jamova 39, 1000 Ljubljana, Slovenia

‡ Electrotechnical Laboratory, 1-1-4 Umezono, Tsukuba, Ibaraki 305–8568, Japan

Organic ferromagnets, which exhibit exchange interactions between unpaired electrons in π -orbitals, are rare^{1–6}, and the origin of ferromagnetism in these compounds has so far remained unexplained. Tetrakis(dimethylamino)ethylene–fullerene[60] (TDAE–C₆₀) shows a transition to a ferromagnetic state with fully saturated $s = 1/2$ molecular spins at the relatively high Curie temperature (for organic materials) of 16 K (ref. 4). It has been suggested^{7–9} that the orientations of the C₆₀ molecules may be important for ferromagnetism in this material, but in the absence of structural data at low temperatures there has been little progress towards understanding these microscopic interactions. Here we report the results of a comparative structural study of two different magnetic forms of TDAE–C₆₀ crystals at low temperatures, correlating the structural properties—in particular, the intermolecular orientations—with the magnetic properties. We find that both ferromagnetism and spin-glass-like ordering are possible in this material, and depend on the orientational state of C₆₀ molecules. This resolves the apparent contradictions posed by different macroscopic measurements^{4,10–14}, and opens the way to a microscopic understanding of π -electron ferromagnetic exchange interactions in organic materials.

The room-temperature structure of ferromagnetic TDAE–C₆₀ crystals is monoclinic (C2/c; ref. 15), with two formula units per unit cell and rapidly rotating C₆₀ molecules¹⁶. As the temperature is lowered, the rotation of the C₆₀ molecules gradually slows down and although NMR data suggest that the rotational motion ‘freezes out’ below about 150 K, the onset of the ferromagnetic (FM) state occurs at a significantly lower temperature; this suggests that stationary C₆₀ molecules may be a necessary but not sufficient condition for the occurrence of a FM state. Fortunately, from the point of view of understanding the low-temperature magnetic interactions, TDAE–C₆₀ has recently been found to exist in two modifications, one FM and one paramagnetic (PM).

Fresh, single crystals of TDAE–C₆₀ when grown below 10 °C (ref. 17) show no FM behaviour and are PM down to 2 K (the α' -phase). On annealing¹⁸ at high temperature they transform into the FM α phase, which displays all the features expected of a proper ferromagnet with fully saturated spins below the Curie temperature, T_C , of 16 K. The two forms appear structurally indistinguishable at room temperature, but we will show here that minute differences between them exist at low temperature, which enable us to identify the molecular configurations associated with the FM phase.

A batch of TDAE–C₆₀ crystals were grown for this study by the diffusion method: a solution of C₆₀ in toluene (2 mg ml^{–1}, Hoechst gold grade) and a 3:1 mixture of toluene and TDAE (Aldrich, 95% pure) were poured into two compartments of a growth cell separated by fritted glass. The cell was then kept at a constant temperature of 8 °C for six months, whereupon crystals were extracted from it. The measurements were performed on a selected sample in the PM phase without annealing, and compared with the same sample after transformation to the FM phase by annealing for 6 h at 70 °C.

The magnetization (M) was measured with a Quantum Design SQUID magnetometer at 2 K, and is shown in Fig. 1a. The magnetization curve of the PM sample is in quantitative agreement with the predicted behaviour for a paramagnet with one spin per formula unit: $M = N\mu_B \tanh[\mu_B H/k_B T]$ (with $N = 1$ spin per mol). Here μ_B is the Bohr magneton, k_B is Boltzmann’s constant, T is temperature and H is the magnetic field. The FM sample on the other hand shows an M – H curve characteristic of proper ferromagnets, reaching full saturation near $H = 50$ Oe. The zero-field a.c. susceptibility of the two samples as measured with an AcSuS susceptometer (System Projekt) is shown in Fig. 1b, the FM sample showing a sharp transition to a FM state at 16 K.

The structural measurements on the two samples were performed using a Weissenberg-type X-ray imaging plate. The structure of both

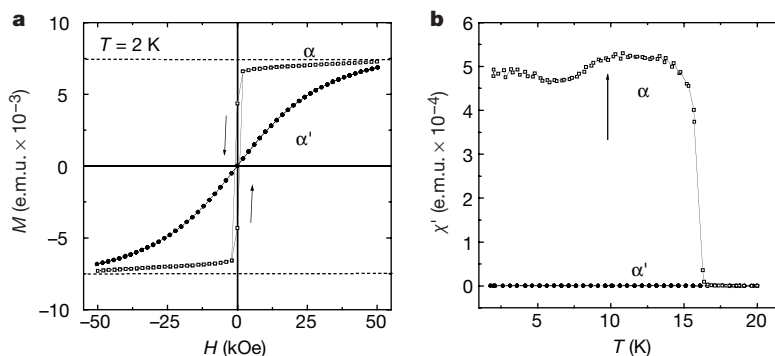


Figure 1 Magnetic behaviour and a.c. susceptibility of PM (α' -phase) and FM (α -phase) samples of TDAE–C₆₀. **a**, The magnetization curves of the PM and FM samples (dots and squares, respectively). The line joining the dots in the PM sample curve is the predicted paramagnetic magnetization $M = N\mu_B \tanh[\mu_B H/k_B T]$ with one spin per formula unit. The FM sample shows a highly nonlinear magnetization M – H curve, characteristic of proper

ferromagnets. The dashed lines correspond to full saturation at $M = N\mu_B$ with $N = 1$ spin per mol. **b**, The a.c. susceptibility χ' of the PM and FM samples. The arrow shows a dip in the susceptibility due to the onset of glass-like behaviour as suggested by the model (see text).

samples at 160 K was found to be in agreement with the established structure¹⁵. The full structural lattice parameters as a function of temperature will be reported elsewhere (B.N., H.K., M.T., A.O. and D.M., unpublished data; see also Supplementary Information); here we concentrate on the parameters relevant for the FM interaction, more specifically the differences in the C_{60} ordering observed between the two phases.

On cooling below 50 K, changes are found in the X-ray diffraction patterns of both PM and FM samples—new diffuse lines appear. In the PM sample these lines disappear in a few hours at low temperature, while in the case of the FM sample they transform into additional sharp diffraction spots. (The data used here were obtained after a 4-h cooling period at 11 K and 7 K on the PM and FM samples, respectively.) In the FM phase, the positions of the additional diffraction spots coincide with those of a primitive unit cell, indicating that the crystal transforms from a C-centred structure to a primitive one. However, our attempts to solve the structure using a primitive unit cell with only single C_{60} orientations resulted in unacceptably high R -factors (about 0.16) and large divergences of the temperature factors for individual atoms. Refinement of the structure in a C-centred unit cell resulted in even worse R -factors (> 0.20), but with large thermal ellipsoids of only some C_{60} atoms, which indicated a high degree of orientational disorder of C_{60} molecules. Finally, using a model for the FM phase with C_{60} molecules rotated relative to each other by 60° about their three-fold axis with 50% relative occupancy (see Fig. 2), the refinement procedure gave a greatly improved R -factor of 0.059, one of the best achieved for any fullerene compound. (A final refinement, where the two occupancies are left unconstrained leads to a I/II occupancy ratio of 53/47%; see below).

Thus, in the FM phase of TDAE- C_{60} , in addition to the conventional 120° rotations found in other C_{60} -based crystal structures¹⁹, the fullerene molecules exist in configurations rotated by $\pm 60^\circ$ about their three-fold molecular axes (Fig. 2). This leads to a set of interfullerene contact configurations that, to our knowledge, have not been previously observed in C_{60} -based solids. As a consequence of these orientations, in the FM phase (and not in the PM phase), the positions of the additional diffraction spots coincide with those of a primitive unit cell, indicating that the crystal transforms from a C-centred structure to a primitive one at low temperature, as shown schematically in Fig. 2.

The relative C_{60} contact configurations corresponding to the new low-temperature orientations, projected along the c axis (which is also the direction of closest contact between the C_{60} s), are shown in Fig. 3. In the PM sample, the relative C_{60} orientations are similar to those encountered in other C_{60} solids: the 6-6 double bond (nearly) faces the centre of the hexagon on the neighbouring molecule (Fig. 3a)¹⁹; we denote this orientation as I. In the FM samples on

the other hand, a new orientation appears (II), which in principle leads to three different possible relative configurations of the C_{60} s, as shown in Fig. 3b–d. However, only one of these configurations—that shown in Fig. 3d—is compatible with the near 50% occupancy of configurations I and II and C-centred primitive unit cell determined from the structural refinement. In this configuration, the double bond on one molecule approximately faces the centre of the pentagon of its neighbour (Fig. 3d), leading to C_{60} molecules ordered along the c axis with alternating orientations I and II that have equal occupancy in the ground state.

We note that the observed arrangement associated with the FM state is in excellent agreement with calculations of the angular dependence of the effective FM exchange coupling strength J_{eff} (ref. 8), which also shows a minimum corresponding to the I-II configuration in Fig. 3d, corresponding to the c -axis Euler angle $\gamma \approx 30^\circ$.

With two equilibrium configurations for the C_{60} molecules, we can describe the magnetic behaviour of the system in terms of a hamiltonian (H) for a non-interacting two-level system in which the coupling between the configurational and magnetic degrees of freedom appears because the exchange interaction J depends on the overlap between adjacent molecules along the c axis:

$$H = - \sum_{ij} J_{ij}(x) \mathbf{S}_i \cdot \mathbf{S}_j + \Delta \sum_i \left(\sigma_i^z + \frac{1}{2} \right)$$

Here Δ describes the energy difference between the two configurations (I-I and I-II in Fig. 3), x is the relative concentration of the configuration favouring a FM exchange interaction, and in general J_{ij} is a function of σ_i^z , where $\mathbf{S}$ is the spin and σ^z is an operator connecting the two configurations. We assume that a spontaneous magnetization appears if the concentration $x = \langle \sigma_i^z + \frac{1}{2} \rangle$ is larger than a certain critical value x_c (this can be understood as a percolative threshold for the appearance of an infinite magnetic cluster). Thus $J_{ij}(x) = J$ if $x > x_c$ and $J_{ij}(x) = 0$ if $x < x_c$. We can write the free energy of the system as the sum of the magnetic

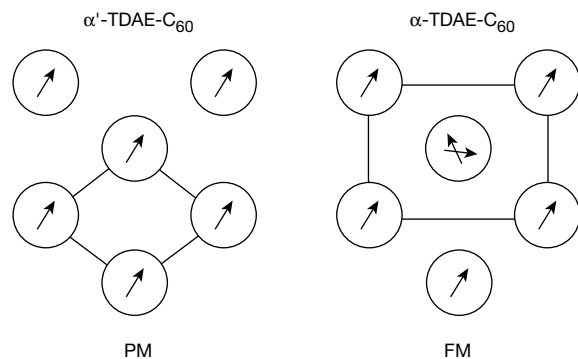


Figure 2 A schematic diagram of the C_{60} molecular orientations in the a - b plane for the PM (left) and FM structures (right). The corresponding C-centred and primitive unit cell bases in the a - b plane are shown.

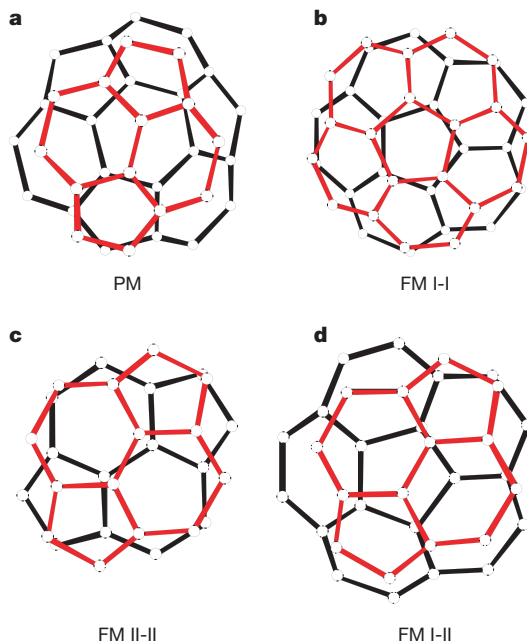


Figure 3 Projections of two neighbouring C_{60} units of TDAE- C_{60} along the direction from C_{60} centre to C_{60} centre in the $[001]$ direction. Upper molecules are shown in red. **a**, The PM phase. **b**, The I-I configuration of C_{60} molecules (FM phase). **c**, The II-II configuration of C_{60} molecules (FM phase). **d**, The I-II configuration (FM phase).

energy, the configurational energy and an entropy term as:

$$F = -kT_c(x)M^2/2 + \Delta x + T[x\ln(x) + (1-x)\ln(1-x)]$$

where M is the magnetization and T_c depends on x via J . If the configurational energy Δ is small (corresponding to the FM phase of TDAE- C_{60}), $T_c > -\Delta/\ln[x_c/(1-x_c)]$, and the magnetic energy exceeds the configurational energy, then a FM magnetic phase is stable below T_c . In the opposite case, (corresponding to the PM phase), the configurational energy is always greater than the magnetic energy, and a ferromagnetically ordered phase is unfavourable. We point out that a situation can arise whereby as T is lowered (below T_c), x can fall below x_c , which means that the magnetic energy can no longer overcome the configurational energy and a first-order transition to a non-magnetic phase can take place at $T < T_c$. When this occurs, we expect that phase separation will take place and non-magnetic clusters will appear. The manifestation of this behaviour is a glass-like or superparamagnetic-like response (depending on the size of the magnetic clusters). This would explain, for example, both the re-entrant spin-glass-like behaviour and the fall in the magnetization (or susceptibility) observed below T_c (ref. 10), as shown in Fig. 1b.

The identification of the C_{60} orientations associated with the FM interaction may lead to the synthesis of a high-temperature fullerene ferromagnet, provided that the I-II configuration can be stabilized by 'fine tuning' of the C_{60} orientations—for example, by the use of C_{60} adducts. More immediately, the data open the way for molecular modelling of the FM exchange interactions, which we hope will soon lead to a better understanding of π -electron FM exchange interactions in organic materials. Although the detailed consequences of the proposed model have not been elucidated here, we have shown that it naturally resolves the controversy raised by a number of seemingly contradictory reports of spin-glass^{8,10}, superparamagnetic¹¹ and ferromagnetic¹⁴ ordering in TDAE- C_{60} , and can predict in detail some of the unique features of this system. To our knowledge there is no other material reported to date which can reversibly display either ferromagnetism or spin-glass properties in response to an externally controllable parameter, such as the cooling rate below 50 K (ref. 7) or the application of a magnetic field below T_c (ref. 4). □

Received 25 October 1999; accepted 15 August 2000.

- Kahn, O. *Molecular Magnetism* (VCH, New York, 1993).
- Nakazawa, Y. *et al.* Low-temperature magnetic properties of the ferromagnetic organic radical, *p*-nitrophenyl nitronyl nitroxide. *Phys. Rev. B* **46**, 8906–8914 (1992).
- Chiarelli, R., Novak, M. A., Rassat, A. & Tholence, J. L. A ferromagnetic transition at 1.48 K in an organic nitroxide. *Nature* **363**, 147–149 (1993).
- Allemand, P. M. *et al.* Organic molecular soft ferromagnetism in a fullerene C_{60} . *Science* **253**, 301–303 (1991).
- Mrzel, A. *et al.* Ferromagnetism in cobaltocene-doped fullerene derivative below 19K due to unpaired spins only on fullerene molecules. *Chem. Phys. Lett.* **298**, 329–333 (1998).
- Umek, P., Omerzu, A., Mihailovic, D. & Tokumoto, M. Synthesis and magnetic characterisation of fullerene derivative-based ferromagnets (3-nitro)- and (3-aminophenyl)-1-H-methanofullerene doped with cobaltocene. *Chem. Phys.* **253**, 361–366 (2000).
- Mihailovic, D. *et al.* Orientational and magnetic ordering of buckyballs in TDAE- C_{60} . *Science* **268**, 400–402 (1995).
- Sato, T., Saito, T., Yamabe, T. & Tanaka, K. Spin-glassy phase of TDAE- C_{60} crystals. *Phys. Rev. B* **55**, 11052–11055 (1997).
- Tanaka, K. *et al.* Orientation dependent magnetic interaction in TDAE- C_{60} where TDAE is tetrakis(dimethylamino)ethylene. *Chem. Phys. Lett.* **259**, 574–579 (1996).
- Tanaka, K. *et al.* Ac magnetic susceptibility of TDAE- C_{60} . *Chem. Phys. Lett.* **237**, 123–126 (1995).
- Tanaka, K. *et al.* Magnetic properties of TDAE- C_{60} and TDAE- C_{70} where TDAE is tetrakis(dimethylamino)ethylene. *Phys. Rev. B* **47**, 7554–7559 (1993).
- Venturini, P. *et al.* Magnetic resonance investigation of the magnetic transition in TDAE- C_{60} . *Int. J. Mod. Phys. B* **6**, 3947–3951 (1992).
- Blin, R., Cevc, P., Arcon, D., Mihailovic, D. & Venturini, P. Time decay of the remanent magnetization in TDAE- C_{60} . *Phys. Rev. B* **50**, 1–3 (1994).
- Arcon, D., Cevc, P., Omerzu, A. & Blin, R. Ferromagnetic resonance in TDAE- C_{60} . *Phys. Rev. Lett.* **80**, 1529–1533 (1998).
- Golic, L. *et al.* in *Fullerenes and Fullerene Nanostructures* (eds Kuzmany, H., Fink, J., Mehring, M. & Roth, S.) 531–534 (World Scientific, Singapore, 1996).
- Arcon, D., Dolinsek, J. & Blin, R. ^{13}C NMR of the organic ferromagnet TDAE- C_{60} . *Phys. Rev. B* **53**, 9137–9142 (1996).

- Omerzu, A., Mihailovic, D. & Tokumoto, M. Macroscopic magnetic measurements on the non-ferromagnetic modification of TDAE- C_{60} . *Phys. Rev. B* **61**, R11883–R11885 (2000).
- Mrzel, A., Cevc, P., Omerzu, A. & Mihailovic, D. Interplay between antiferromagnetic and ferromagnetic phases of TDAE- C_{60} : An ESR study of high-temperature annealed samples. *Phys. Rev. B* **53**, R2922–R2925 (1996).
- Dresselhaus, M. S., Dresselhaus, G. & Eklund, P. C. *Science of Fullerenes and Carbon Nanotubes* (Academic, San Diego, 1996).

Supplementary Information is available on Nature's World-Wide Web site (<http://www.nature.com>) or as paper copy from the London editorial office of Nature.

Correspondence and requests for materials should be addressed to D.M. (e-mail: dragan.mihailovic@ijs.si).

Electrochemically induced annealing of stainless-steel surfaces

G. T. Burstein, I. M. Hutchings & K. Sasaki

Department of Materials Science and Metallurgy, University of Cambridge, Pembroke Street, Cambridge CB2 3QZ, UK

Modification of the surface properties of metals without affecting their bulk properties is of technological interest in demanding applications where surface stability and hardness are important. When austenitic stainless steel is heavily plastically deformed by grinding or rolling, a martensitic phase transformation occurs that causes significant changes in the bulk and surface mechanical properties^{1–9} of the alloy. This martensitic phase can also be generated in stainless-steel surfaces by cathodic charging, as a consequence of lattice strain generated by absorbed hydrogen^{10–16}. Heat treatment of the steel to temperatures of several hundred degrees can result in loss of the martensitic structure¹, but this alters the bulk properties of the alloy. Here we show that martensitic structures in stainless steel can be removed by appropriate electrochemical treatment in aqueous solutions at much lower temperature than conventional annealing treatments. This electrochemically induced annealing process allows the hardness of cold-worked stainless steels to be maintained, while eliminating the brittle martensitic phase from the surface. Using this approach, we are able to anneal the surface and near-surface regions of specimens that contain rolling-induced martensite throughout their bulk, as well as those containing surface martensite induced by grinding. Although the origin of the electrochemical annealing process still needs further clarification, we expect that this treatment will lead to further development in enhancing the surface properties of metals.

Samples of an austenitic sheet stainless steel—containing 18% chromium, 8% nickel and 0.03% carbon (AISI type 304L)—were cut into 1-cm squares, each carrying a thin shank from one corner to allow electrical contact. The surfaces of each were ground and polished to a 0.25- μm finish with successive grades of silicon carbide paper and diamond paste. X-ray diffraction of this surface showed the presence of the austenitic structure (γ) together with reflections from the martensitic structure (α'); see Fig. 1a. The martensitic structure, which is generated by the strain developed during mechanical surface preparation, has its most prominent reflection at $2\theta = 44.69^\circ$, representing the (110) reflection¹⁰.

The surface was then subjected to a series of anodic/cathodic electrochemical pulses in an 8 M aqueous solution of sodium nitrite at a temperature of 80 °C. This was carried out using a two-electrode cell with cell voltages alternately of -4.1 V and $+1.2\text{ V}$; this is equivalent to applied electrode potential pulses between about -1.8 V (w.r.t. SCE) and about $+0.28\text{ V}$ (w.r.t. SCE)—here w.r.t. SCE

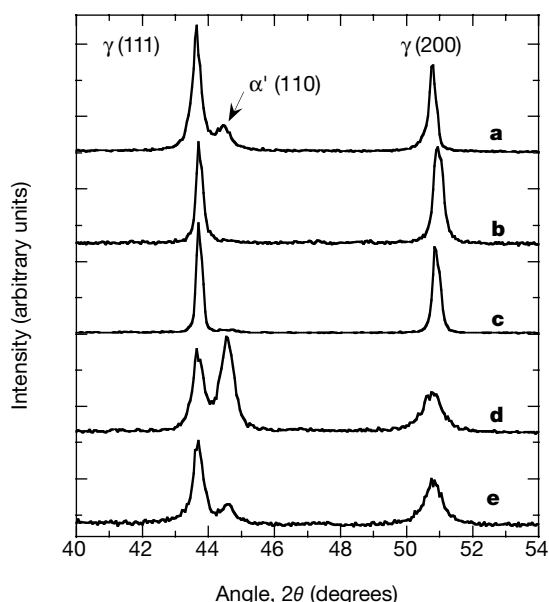


Figure 1 X-ray diffraction patterns of type 304L stainless-steel surfaces with Cu K α radiation. The original surface finish was achieved by polishing to 0.25 μm (see text). The predominant, austenite (γ) phase is shown by its (111) and (200) reflections. **a**, Original surface before any treatment. The martensite phase (α') is shown by its (110) reflection at $2\theta = 44.69^\circ$. **b**, The surface after electrochemical pulse treatment in 8 M sodium nitrite at a temperature of 80 $^\circ\text{C}$. The voltage pulse across the cell was 33 cycles each of -4.1 V for 277 s (cathodic) and $+1.2$ V for 56 s (anodic). The martensite phase has been removed. **c**, The original polished surface after a thermal annealing procedure at 750 $^\circ\text{C}$ for 20 min also shows the loss of the martensite phase. (We note that a very small martensite peak is still visible.) **d**, The steel after cold-rolling to achieve a 25% reduction in thickness shows a large peak due to (110) reflection from α' martensite. **e**, The cold-rolled steel after the same electrochemical treatment as in **b**, but this time for 98 cycles. The peak due to martensite is substantially reduced.

indicates with respect to the standard calomel electrode. The electrode potentials were measured in some experiments using a reference electrode mounted into a tube with a Luggin capillary tip near the specimen surface. The counter electrode was a graphite cylinder. After 3 h of electrochemical treatment, the specimen was withdrawn and washed thoroughly. The resulting X-ray diffraction pattern showed that the reflections due to martensite had disappeared completely (Fig. 1b). The austenite peaks were unchanged. We call this loss of the martensite phase 'electrochemically induced annealing'. The same annealing effect could be achieved by heat treatment in the absence of the electrochemical process (Fig. 1c), although temperatures in excess of 750 $^\circ\text{C}$ were required to achieve this effect in a reasonable time. We attempted to anneal the steel surface electrochemically by using only individual anodic or cathodic treatments, using longer times than taken for the pulses (for example, 6 h cathodic treatment), but no loss of martensite was observed. Clearly the electrochemical annealing was a consequence of the application of both the anodic and cathodic potential pulses, rather than the individual anodic or cathodic parts of the electrochemical signal alone.

The electrochemical annealing process described above gives rise to mechanical properties of the surfaces that are different from those achieved by thermal annealing. We show this in Fig. 2 by presenting the hardness of the surfaces measured using a Vickers microhardness indenter (with a 50-g load). After grinding, the stainless steel surface had an initial hardness (H_V) of 205 kgf mm^{-2} , plotted as bar a in Fig. 2. Thermal annealing at 750 $^\circ\text{C}$ reduces H_V to 196 kgf mm^{-2} , consistent with the loss of strain-induced martensite (Fig. 2, bar b). The thermally annealed specimen was again hardened by regrinding, as shown in Fig. 2, bar c. After the electrochemical annealing procedure described above the hardness was in contrast still high

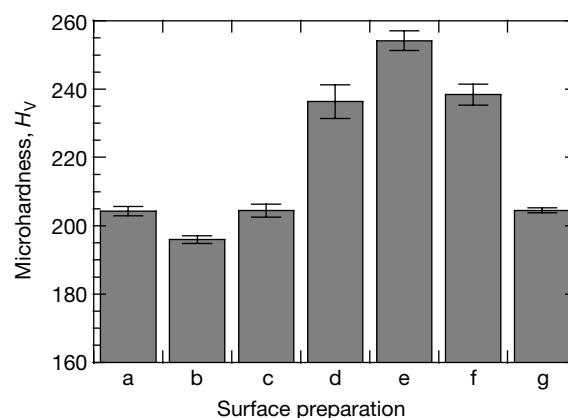


Figure 2 The microhardness of the stainless-steel surfaces with an initial surface finish achieved by polishing to 0.25 μm . Measurements were made using a Vickers diamond indenter with a 50-g load. The scatter band at the top of each bar shows the standard deviation of four or five measurements. **a**, The surface as polished. **b**, After annealing thermally at 750 $^\circ\text{C}$ for 20 min. The hardness has decreased, consistent with the loss of martensite. **c**, The hardness of the annealed surface shown in **b** increases again after polishing to a finish similar to **a**. **d**, The polished surface after electrochemical treatment in 8 M sodium nitrite solution at 80 $^\circ\text{C}$ using 33 cycles of -4.1 V for 277 s (cathodic) and $+1.2$ V for 56 s (anodic) (total treatment time 3 h). The surface is significantly harder despite the loss of the martensite phase. **e**, The same electrochemically treated surface as in **d**, but after a further thermal treatment at 550 $^\circ\text{C}$ for 20 min. The surface has not softened. **f**, The polished surface after an exclusively cathodic electrochemical treatment at -4.5 V for 6 h. The surface has hardened, but martensite was not removed by this treatment. **g**, The same surface as in **f** but after further thermal treatment at 550 $^\circ\text{C}$ for 20 min. The surface hardness has fallen to that of the original polished surface.

($H_V = 235 \text{ kgf mm}^{-2}$), despite the loss of the martensite (Fig. 2, bar d). Thus the surface retains a 'memory' of the fact that strain-induced martensite had been formed and removed electrochemically. Moreover, when the surface had been electrochemically annealed and subsequently thermally annealed at the lower temperature of 550 $^\circ\text{C}$, the hardness remained high ($H_V = 252 \text{ kgf mm}^{-2}$), significantly higher than what was achieved by thermal annealing alone: this is shown in Fig. 2, bar e. The hardness resulting from the electrochemical pulse treatment might at first glance be thought to be induced by the penetration of hydrogen into the metal matrix: this possible explanation was examined experimentally by applying a cathodic treatment alone, without the anodic pulse, using a more negative voltage of -4.5 V and a far longer time of 6 h. Under these conditions, hydrogen is generated cathodically by reduction of water. We note, in comparison, that the total cathodic component of the annealing pulse treatment was only 2.5 h. The hardness generated under this more severe exclusively cathodic treatment (see Fig. 2, bar f) was however removed completely by the subsequent 550 $^\circ\text{C}$ heat-treatment (Fig. 2, bar g), in contrast to the corresponding hardness after pulse treatment (which remained unchanged after treatment at 550 $^\circ\text{C}$). We conclude that the electrochemical annealing procedure engendered by pulse treatment induces relatively irreversible and stable changes into the stainless-steel surfaces involving both loss of the martensite phase and retention of surface hardening.

Finally, because the original 0.25- μm mechanically polished surface finish is a very fine one, we electrochemically annealed specimens of the same steel in which martensite had been induced throughout the structure by cold-rolling. In these experiments, specimens of the same material were cold-rolled to produce a 25% reduction in thickness. The X-ray diffraction pattern displayed in Fig. 1d shows a pronounced peak from the martensitic structure. The far greater amplitude of the martensite peak in this case is due to the fact that the rolling treatment generates strain-induced martensite throughout the structure. Electrochemical pulse treat-

ment in the same manner as described above shows substantial loss of the martensite peak over a period of 9 h (after which the experiments were terminated), with no change in the height of the austenite peak (Fig. 1e). Although the procedure took longer to reduce the martensite peak substantially, we nevertheless observed a significant reduction, even after 3 h. The electrochemically induced removal of martensite clearly occurs to some significant depth beneath the metal surface, and is not confined to the surface atoms only. We estimate the depth of annealing after 9-h electrochemical treatment to be at least 8 μm .

Although the origins of electrochemical annealing are not yet clear, some aspects can be rationalized. In the cathodic part of the pulse, water is reduced by electrolysis to hydrogen. Hydrogen atoms entering the metal produce considerable strain within the lattice, causing microstructural changes as well as phase transformations. Previous work on cathodic charging gives conflicting information. Cathodic charging is reported to generate both the α' as well as the hexagonal ϵ phases of martensite^{10,12}. In contrast to this, it is also reported that only the α' phase, and not the ϵ phase, is generated for the same grade of stainless steel¹⁴. Okada *et al.*¹² show that ϵ and α' martensite are formed during cathodic charging of type 304 stainless steel, both of which persist after ageing, although type 310 stainless steel (a more highly alloyed material containing 20% Cr and 22% Ni) showed ϵ martensite during cathodic charging which subsequently decomposed during ageing. Our observations from pulse experiments differ substantially. The pulsed annealing treatment destroys the α' martensite phase, rather than generating it. No sign of ϵ phase (whose largest reflection would be observed¹⁰ at $2\theta = 46.96^\circ$; Fig. 1) was found, either during mechanical deformation, or after electrochemical treatment. The pulse treatment leaves the structure entirely austenitic, a phenomenon not before observed, to our knowledge. Kinetically, the hydrogen reaction could account for the observed changes. By assuming the rate-determining step to be diffusion within the metal phase, the minimum annealed depth of 8 μm estimated above gives a diffusion coefficient of about $2 \times 10^{-11} \text{ cm}^2 \text{ s}^{-1}$. This corresponds quite well to the value of $10^{-10} \text{ cm}^2 \text{ s}^{-1}$ for thermally annealed 304 stainless steel obtained by extrapolating from high-temperature data¹⁷ to our experimental temperature of 80 °C. However, one cannot preclude possible synergistic effects of nitrogen produced by cathodic reduction of the electrolytic nitrite ion. Examination of the treated metal by energy dispersive X-ray analysis in a scanning electron microscope, and by energy loss spectroscopy in the transmission electron microscope (both methods *ex situ*) failed to reveal the presence of nitrogen. Nevertheless, we were unable to achieve electrochemical annealing in aqueous potassium carbonate solution using similar electrochemical surface treatment. Nitrite is certainly electrochemically reducible, at least to ammonia¹⁸, and reduction to atomic nitrogen is in principle possible.

In the anodic part of the potential pulse, the metal passivates by oxide film growth, in accordance with the classical behaviour of stainless steels. (Transpassive dissolution of the chromium component does not occur at the potentials imposed here.) Any reduced hydrogen and nitrogen generated in the previous cathodic pulse and dissolved into the metal matrix would also tend to be re-oxidized anodically and dissolve into the electrolyte. The defects generated by ingress of hydrogen (and perhaps nitrogen) would leave vacancies after re-oxidation and provide for surface relaxation of the metal atoms over very short range, giving the observed annealing. The defects that remain would then be available for penetration of further reduced atoms in the subsequent cathodic pulse. In this way a deeply affected layer could be induced. To our knowledge, this is the first time that such an annealing phase change procured by electrochemical means has been reported.

The present work paves the way for further development of new methods of electrochemical surface processing that could enhance the surface properties of metals, without altering their

bulk properties—in contrast to the action of heat treatment, which acts on both the surface and the bulk. □

Received 30 June; accepted 14 August 2000.

- Sumitomo, H. Investigation of the development of cold-rolling and annealing textures in SUS304 austenitic stainless steel. *J. Iron Steel Inst. Jpn* **77**, 558–565 (1991).
- Angel, T. Formation of martensite in austenitic stainless steel. *J. Iron Steel Inst.* **177**, 165–174 (1954).
- Ludwigson, D. C. & Berger, J. A. Plastic behaviour of metastable austenitic stainless steels. *J. Iron Steel Inst.* **207**, 63–69 (1969).
- Olson, G. B. & Cohen, M. Kinetics of strain-induced martensitic transformation. *Metall. Trans. A* **6**, 791–795 (1975).
- Olson, G. B. & Cohen, M. Stress-assisted isothermal martensitic transformation: application to TRIP steels. *Metall. Trans. A* **13**, 1907–1914 (1982).
- Huang, G. L., Matlock, D. K. & Krauss, G. Martensite formation, strain rate sensitivity and deformation behaviour of type 304 stainless steel sheet. *Metall. Trans. A* **20**, 1239–1246 (1989).
- Tomimura, K., Takaki, S. & Tokunaga, Y. Reversion mechanism from deformation induced martensite to austenite in metastable austenitic stainless steels. *ISIJ Int.* **31**, 1431–1437 (1991).
- Takaki, S., Tomimura, K. & Ueda, S. Effect of pre-cold-working on diffusional reversion of deformation induced martensite in metastable austenitic stainless steel. *ISIJ Int.* **34**, 522–527 (1994).
- Petersen, S. F., Mataya, M. C. & Matlock, D. K. The formability of austenitic stainless steels. *J. Mater.* **49**, 54–58 (1997).
- Bentley, A. P. & Smith, G. C. Phase transformation of austenitic stainless steels as a result of cathodic hydrogen charging. *Metall. Trans. A* **17**, 1593–1600 (1986).
- Tähtinen, S., Kivilahti, J. & Hanninen, H. Crystallography of hydrogen-induced surface cracking in a spheroidal austenitic stainless steel single crystal. *Scripta Metallurgica* **19**, 967–973 (1985).
- Okada, H., Hosoi, Y. & Abe, S. Formation of cracks in austenitic stainless steels cathodically charged with hydrogen. *Corrosion* **26**, 183–186 (1970).
- Ohtani, N., Asano, S., Fujishima, Y. & Yamamasu, Y. Hydrogen-induced transformation and embrittlement in 18-8 stainless steel. *J. Jpn Inst. Metals* **37**, 746–753 (1973).
- Narita, N., Altstetter, C. J. & Birnbaum, H. K. Hydrogen-related phase transformations in austenitic stainless steels. *Metall. Trans. A* **13**, 1355–1365 (1982).
- Szumner, A. & Janko, A. Hydride phases in austenitic stainless steels. *Corrosion* **35**, 461–464 (1979).
- Szumner, A. in *Hydrogen Degradation of Ferrous Metals* (eds Oriani, R. A., Hirth, J. P. & Smalowski, M.) 512–534 (Noyes Publications, Park Ridge, New Jersey, 1985).
- Perng, T. P. & Altstetter, C. J. Effects of deformation on hydrogen permeation in austenitic stainless steels. *Acta Metall.* **34**, 1771–1781 (1986).
- Newman, R. C. & Burstein, G. T. Anion effects in the stress-corrosion cracking of copper and brass. *J. Electrochem. Soc.* **127**, 2527 (1980).

Acknowledgements

We thank P.A. Midgley and M. Weyland for the TEM analysis. This work was supported by the EPSRC under the ROPA scheme.

Correspondence and requests for materials should be addressed to G.T.B. (e-mail: gtb1000@cus.cam.ac.uk).

Cooler winters as a possible cause of mass extinctions at the Eocene/Oligocene boundary

Linda C. Ivany*†, William P. Patterson†‡ & Kyger C Lohmann*

* Department of Geological Sciences, University of Michigan, Ann Arbor, Michigan 48109, USA

‡ Department of Earth Sciences, Syracuse University, Syracuse, New York 13244, USA

The Eocene/Oligocene boundary, at about 33.7 Myr ago, marks one of the largest extinctions of marine invertebrates in the Cenozoic period¹. For example, turnover of mollusc species in the US Gulf coastal plain was over 90% at this time^{2,3}. A temperature change across this boundary—from warm Eocene climates to cooler conditions in the Oligocene—has been suggested as a cause of this extinction event⁴, but climate reconstructions have not provided support for this hypothesis. Here we report stable oxygen isotope measurements of aragonite in fish otoliths—ear stones—collected across the Eocene/Oligocene boundary. Palaeotemperatures reconstructed from mean otolith oxygen isotope

† Present address: Department of Earth Sciences, Syracuse University, Syracuse, New York, 13244, USA.

values show little change through this interval, in agreement with previous studies^{5,6}. From incremental microsampling of otoliths, however, we can resolve the seasonal variation in temperature, recorded as the otoliths continue to accrete new material over the life of the fish. These seasonal data suggest that winters became about 4 °C colder across the Eocene/Oligocene boundary. We suggest that temperature variability, rather than change in mean annual temperature, helped to cause faunal turnover during this transition.

The US Gulf coastal plain preserves one of the best-studied Palaeogene marine shelf sections in the world. The molluscan fossil record is especially well constrained owing to more than a century of intensive collecting and the compilation of a standardized taxonomy^{7–10}. Turnover at the Eocene/Oligocene (E/O) boundary is very high, with fewer than 10% of late Eocene species surviving into the Oligocene^{2,3}. Although temperature remains the most likely cause of this turnover^{4,11}, no studies to date have been able to demonstrate conclusively a temperature change specific to the boundary in this region. Temperatures derived from the $\delta^{18}\text{O}$ of planktonic foraminifers from the Gulf of Mexico, the Caribbean, and low-latitude North Atlantic Ocean show no change in temperature across the boundary⁵. Palynomorph (pollen and spores) assemblages from Gulf Coast sections suggest a progressive cooling and drying on land beginning in the late Eocene and continuing into the early Oligocene^{6,12}, but detailed palaeofloristic work could resolve no significant climate change at the boundary⁶. Conversely, calcareous nanoplankton assemblages suggest that cooling began only later in the early Oligocene, after the boundary extinctions¹³. If temperature is to remain a viable mechanism for extinction, it is necessary that data on temperature show some excursion at the boundary.

To test the hypothesis that temperature played a role in faunal turnover, and to shed light more broadly on environmental changes that may have occurred on the shelf during this interval of time, we sectioned and incrementally microsampled fish otoliths collected

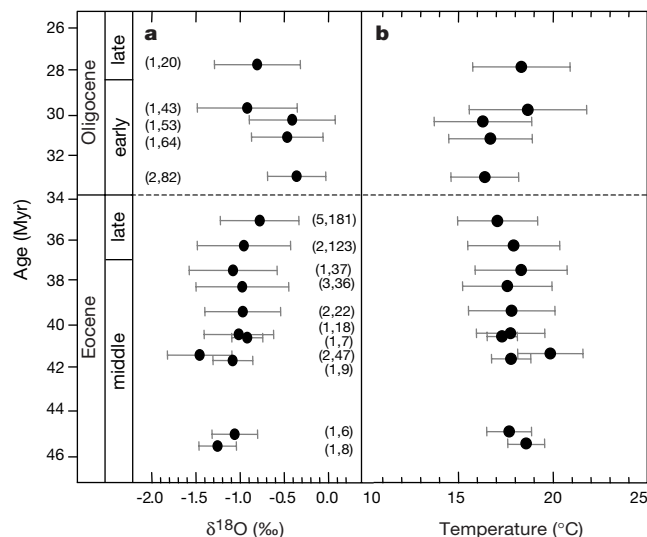


Figure 1 Results from isotopic analyses of microsampled otoliths. The figure shows mean annual $\delta^{18}\text{O}$ (a) and mean annual temperature (b) for the middle Eocene to the end of the Oligocene. Values plotted are means of all analyses from all otoliths from a given stratigraphic horizon, including both congrid and ophidiids; error bars are 1 standard deviation around the mean. Number of otoliths and total number of analyses per interval are given in parentheses; a total of 756 isotopic analyses on microsamples from 26 otoliths are reported. See text for values of Palaeogene seawater $\delta^{18}\text{O}$ used in temperature calculations. Timescale is from ref. 28. Ages of samples are the mean age of the formation/member from which they came, as interpolated from planktonic foraminifer zones given in refs 29 and 30. All $\delta^{18}\text{O}$ values are reported relative to VPDB.

across the E/O boundary in the Gulf Coast. Stable oxygen isotopic compositions of aragonite subsamples provide a proxy record for temperature during the lifetime of fish^{14–17}. Moreover, high-resolution analyses of accretionary otolith aragonite enable the approximation not only of mean annual temperature but also of the seasonal range of temperature variation. This technique provides an advantage over more traditional proxies of bulk carbonate or whole-shell (for example, foraminiferal) analyses because seasonality is a factor inaccessible from them but crucial in determining the biogeographic distributions of organisms today^{18,19}.

Mean $\delta^{18}\text{O}$ values of congrid and ophidiid otoliths (Fig. 1a) increase slightly ($\sim 0.4\text{‰}$) at the middle/late Eocene boundary and again by $\sim 0.6\text{‰}$ across the E/O boundary. Similar increases (although of a somewhat higher magnitude across the E/O boundary) have been observed in numerous records of benthic foraminifer $\delta^{18}\text{O}$ in the deep ocean^{5,20–22}. However, when mean palaeotemperatures are calculated (Fig. 1b), the data reveal little change from the middle Eocene into the Oligocene. A t -test between middle Eocene and late Eocene samples grouped by age is statistically insignificant (0.3 °C , $P = 0.613$, $n = 20$ otoliths). The difference between late Eocene and early Oligocene mean temperatures (1.1 °C) is also statistically insignificant ($P = 0.186$, $n = 12$ otoliths). The maximal

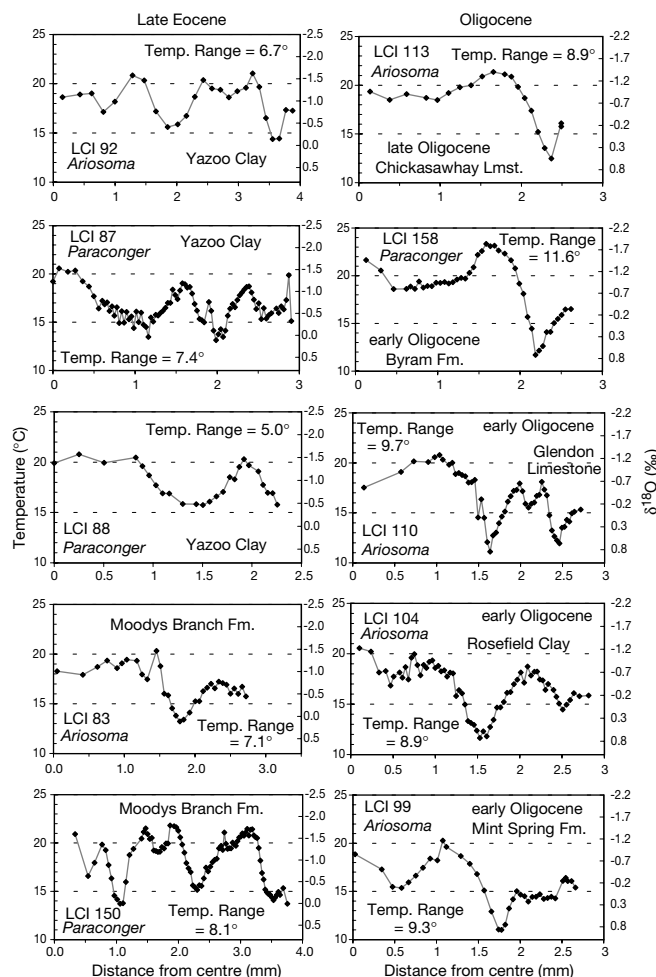


Figure 2 Individual temperature records from microsampled congrid otoliths. Records are shown from the late Eocene (left) and the early and late Oligocene (right). Sample number, genus, formation and seasonal range of temperature for each otolith are indicated on the plots. In each panel, temperature and $\delta^{18}\text{O}$ are shown on the left- and right-hand vertical axis, respectively. We note that the early Oligocene is more seasonal and has cooler winters than the late Eocene, yet summer temperatures do not differ appreciably.

change across the Eocene/Oligocene boundary as indicated by a comparison of latest Eocene and earliest Oligocene samples is 1.8 °C, but again is statistically insignificant ($P = 0.202$, $n = 7$ otoliths). The lack of substantial change in mean temperature across the Eocene/Oligocene boundary is consistent with earlier work in the Gulf Coast and with Mg/Ca ratios of deep-sea benthic foraminifers²³. Relative consistency in temperature through time might be expected in the Gulf Coast because cooling during the transition from early Eocene warmth to Oligocene glacial conditions was far more extreme at high latitudes than at low⁵. Our data suggest that if there was cooling in the Gulf from an early Eocene thermal maximum, the bulk of it must have occurred before the middle of the middle Eocene.

Beyond estimates of mean annual temperature, $\delta^{18}\text{O}$ values of individual serially microsampled otoliths also record the seasonal temperature variation experienced throughout the lifetimes of the fish. Figure 2 shows temperature profiles from ten late Eocene and early Oligocene congrid otoliths. The range of temperatures experienced in the late Eocene generally falls between 15 and 20 °C, with winter minima not extending below about 13.5 °C. The Oligocene congrid records record similar warm month temperatures of around 20 °C, but winter temperatures consistently fall below those of the Eocene samples, with values as low as 11.0 °C. The scalloped nature of Oligocene curves also suggests that otolith growth may have slowed significantly or stopped during the coldest month(s)²⁴, indicating that winter temperatures may have been even colder than those recorded in the otolith carbonate. The trend in minimum and maximum temperatures from the late Eocene into the Oligocene is shown in Fig. 3a. A t -test of summer and winter temperatures recorded by late Eocene and Oligocene otoliths demonstrates no significant change in summer temperatures ($P = 0.800$, $n = 21$ summers in 12 otoliths), while Oligocene winters were colder than late Eocene winters ($P = 0.025$, $n = 21$ winters in 12 otoliths) by as much as 4 °C. Figure 3b compares the seasonal range of temperature variation in the late Eocene versus the Oligocene, and shows that seasonality increases significantly across the boundary.

These data suggest that while summer temperatures did not change appreciably over the Eocene/Oligocene transition, cold

month mean temperatures dropped by some 4 °C, or more if cold month extremes are not actually recorded. Such a decrease in winter minimum temperature on the shelf would have a significant effect on the biogeographic ranges of taxa, and could be responsible for the extinctions of stenothermal (unable to tolerate large fluctuations in temperature) tropical taxa in the Gulf. The pattern of a drop in winter temperatures without a concomitant change in summer temperatures may also help to discriminate between alternative explanations for Eocene/Oligocene cooling. A decrease in the concentration of greenhouse gases may be more likely to result in uniform degrees of cooling in summer and winter months, while changes in oceanographic and/or atmospheric circulation could produce more variable results.

Several caveats about these data and the conclusions drawn from them should be mentioned. First and perhaps most serious is the question of how representative of an interval of time encompassed by an individual sampling horizon (of the order of 10^3 – 10^4 years) is the record of seasonal variation preserved in one or several otoliths, from fish that lived for 2–3 years. Although this is a potential limitation, samples from different otoliths taken from different formations at different localities through the late Eocene and early Oligocene show a generally consistent pattern in the temperatures they generate; hence, the trend appears to be robust.

Other potential limitations are associated with interpreting seasonal ranges of temperature from discretely sampled carbonate. By necessity, the degree of annual temperature variation will correlate with sampling resolution. The more finely resolved the subsampling, the greater the range in inferred temperature; conversely, the coarser the sampling, the greater the amount of time averaged within each sample, and the more seasonal amplitudes will be reduced. As mentioned above, this may be a concern with the ophidiid otoliths. Sampling resolution does vary to some degree among the late Eocene and Oligocene congrid otoliths presented here, yet sampling intensity is always high enough to record a significant number of points from each seasonal cycle. As a result, it is unlikely that inferred seasonal ranges of variation are compromised.

Our data reflect temperature variations experienced at the bottom on a marine shelf. The palaeodepths of these samples are poorly constrained; however, the seasonal range of temperature variation in the northern Gulf at 50 m depth today is about 5 °C (ref. 25), roughly that observed in our data. Temperatures recorded at the sea bottom on the Atlantic shelf in 50 m water depth accurately reflect surface temperatures from November to May, when the water column is isothermic²⁶. Stratification during the summer months prevents maximum bottom-water temperatures from reaching those at the surface, but turnover in the autumn mixes warm surface water to the bottom. This suggests that while our warmest bottom temperatures may underestimate those reached at the surface, cold-month temperatures accurately record surface temperatures as long as Gulf waters experienced turnover.

Our estimates of palaeotemperature are based on assumptions about the $\delta^{18}\text{O}$ of the water from which the otoliths were precipitated (see Methods). Lear *et al.*²³ provide an alternative interpretation of $\delta^{18}\text{O}$ of sea water through time, and suggest that the shift in $\delta^{18}\text{O}$ of carbonate across the Eocene/Oligocene boundary is entirely due to changes in ice volume and hence seawater $\delta^{18}\text{O}$. Using their values for seawater composition inverts our results, suggesting that winter temperatures remained constant while summers became warmer. This is not consistent with the selective extinction of warm-water taxa, and hence we prefer the Zachos *et al.*⁵ estimates of $\delta^{18}\text{O}$ of sea water. It is conceivable that isotopically more negative freshwater runoff may have influenced the $\delta^{18}\text{O}$ of sea water in the early Oligocene due to regression and increasing proximity of the shoreline. However, the observed shift in $\delta^{18}\text{O}$ of carbonate is positive, the opposite direction to what would be expected from a salinity effect, and is consistent with the offset seen in the deep sea

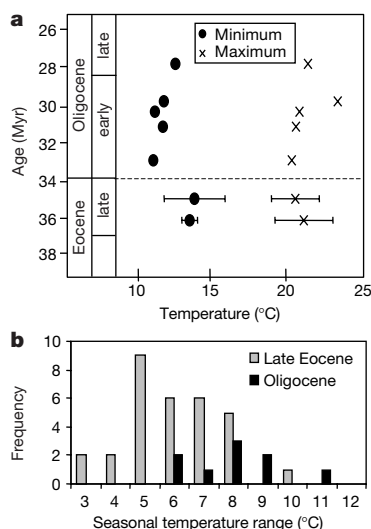


Figure 3 Eocene/Oligocene seasonal temperature change, as inferred from $\delta^{18}\text{O}$ of otoliths. **a**, Minimum and maximum temperatures for late Eocene and Oligocene otoliths from given sampling horizons. **b**, Frequency distribution of seasonal temperature ranges recorded by late Eocene and Oligocene otoliths. Values for range include every combination of local maximum minus local minimum temperature that could be unambiguously recognized as a seasonal peak within each otolith. The difference between late Eocene and Oligocene seasonal temperature range is highly significant ($P = 0.004$, $n = 40$ differences between summer and winter temperatures in 12 otoliths).

record²². Any damping of the magnitude of this shift by freshwater admixture would serve to decrease our estimate of mean temperature change across the boundary, and hence our values should be taken as minimum estimates of change. Nevertheless, consistency with earlier climate work in the Gulf and the lack of palaeontological evidence for reduced salinity suggest that any deviation from normal marine sea water was minimal if present at all. We emphasize that variations in $\delta^{18}\text{O}$ of sea water through time due to local changes in salinity or global changes in ice volume will affect only the absolute values of inferred palaeotemperatures, and will have no effect on our conclusions about the seasonal range of temperature variability.

The well-documented positive shift of roughly 1‰ in $\delta^{18}\text{O}$ of deep ocean foraminifers in the earliest Oligocene^{21,22} must record some combination of cooling and increasing ice volume near the Eocene/Oligocene boundary. That this shift is also manifest, if to a lesser degree, in the lower-latitude shelf waters of the Gulf Coast suggests that increasing ice volume and concomitant change in the $\delta^{18}\text{O}$ of the global ocean is more likely than temperature change alone. Minimal change in mean annual temperature across the boundary, as shown here, supports this contention. By the mid-late Oligocene, isotopic compositions in the Gulf once again approximated those of the late Eocene, suggesting either that the ice sheet was not long-lived and/or that temperatures warmed into the late Oligocene.

With the ability to reconstruct seasonal maximum and minimum temperatures, we find that warm-month temperatures do not change significantly across the Eocene/Oligocene boundary, but that early Oligocene winters were markedly cooler than those in the late Eocene. The drop in cold-month mean temperature and the associated increase in seasonality, rather than a drop in mean annual temperature, probably had a significant effect on the distributions of marine organisms in the Gulf, and may be the primary cause of the large faunal turnover associated with the boundary. □

Methods

Otoliths were collected from middle Eocene to Oligocene fossiliferous, glauconitic, marly sands deposited on an open marine shelf and exposed in western Alabama, Mississippi, Louisiana and east Texas. Additional samples were obtained from D. Nolf at the Institut Royal des Sciences Naturelles de Belgique in Brussels. To reconstruct Palaeogene shelf temperatures, we use otoliths from fish in the families Congridae (conger eels), and Ophidiidae (cusk eels). Modern adults in these groups are fully marine, crypto-benthic in relatively deep water, and non-migratory. All are desirable traits for recording temperatures experienced by the benthic shelf biota from which most of the fossil record comes. Modern congrids have a planktonic larval stage, but a clear signal of surface water conditions was difficult to detect in the data, suggesting little if any effect on resulting patterns. Species belonging to extant genera (*Ariosoma* and *Paraconger* in the congrids, *Lepophidium* in the ophidiids) were chosen to minimize concerns about assumptions of ecological consistency through time.

Otoliths were polished flat on one side, mounted on glass slides, and polished on the other to produce a thin layer through the centre. The outer edge, centre, and any obvious growth bands were digitized, and interpolated growth-line-parallel sampling paths calculated so that sample masses exceed roughly 20 µg. Micromilling was done on a computer-controlled three-dimensional positioning stage set under a fixed high-precision dental drill. Stable oxygen and carbon isotopic compositions of resulting powders were analysed on the Finnigan MAT 252 mass spectrometer at Syracuse University or the Finnigan MAT 251 at the University of Michigan, both coupled with automated carbonate preparation systems. This sampling methodology provides a high-resolution profile of isotopic variation recorded by the otolith throughout the lifetime of the animal, with individual microsamples representing as little as 1.5–2 weeks of growth.

Palaeotemperatures are calculated from the stable oxygen isotopic compositions of otolith aragonite using the experimentally determined otolith palaeotemperature equation in ref. 17. This equation is very close to thermodynamic equilibrium, the slope is indistinguishable from that given in ref. 27 for bivalve aragonite, and it describes data from freshwater and marine fishes equally well: hence we believe it to be robust. We use middle and late Eocene and Oligocene seawater $\delta^{18}\text{O}$ values of −0.9‰, −0.8‰ and −0.5‰, respectively (all relative to VSMOW; ref. 5; composition on an ice-free Earth is −1.0‰). The mean value of all analyses from all otoliths of a given age provides an approximation of mean annual temperature for that time. Calculated mean temperatures may be biased slightly toward temperatures in the season of fastest growth, but this offset should be consistent through time.

Palaeoseasonality can be estimated through a comparison of mean minimum and maximum temperature values for each time interval, and can be assessed with confidence for late Eocene and Oligocene congrid otoliths. Middle Eocene samples are dominated by

ophidiids, and while co-occurring congrids and ophidiids show equivalent mean $\delta^{18}\text{O}$ values, lower otolith precipitation rate by the ophidiids may increase time-averaging and lead to an underestimate of seasonal range. Reliable estimates of middle Eocene seasonal range are therefore unavailable at present. Calculated palaeotemperatures depend on the $\delta^{18}\text{O}$ value assumed for sea water. With revisions of these assumptions, absolute values for temperature will change, but the range of temperature variation within samples will not be affected.

Received 25 February; accepted 23 August 2000.

- Raup, D. M. & Sepkoski, J. J. Periodic extinction of families and genera. *Science* **231**, 833–836 (1986).
- Dockery, D. T. Punctuated succession of Paleogene mollusks in the Northern Gulf Coastal Plain. *Palaaios* **1**, 582–589 (1986).
- Haas, D. M. & Hansen, T. A. Timing of latest Eocene molluscan extinction patterns in Mississippi. *Palaaios* **11**, 487–494 (1996).
- Hansen, T. A. Extinction of Late Eocene to Oligocene molluscs: Relationship to shelf area, temperature changes, and impact events. *Palaaios* **2**, 69–75 (1987).
- Zachos, J. C., Stott, L. D. & Lohmann, K. C. Evolution of early Cenozoic marine temperatures. *Paleoceanography* **9**, 353–387 (1994).
- Oboh, F. E., Jaramillo, C. A. & Reeves-Morris, L. M. Late Eocene-Early Oligocene paleofloristic patterns in southern Mississippi and Alabama, US Gulf Coast. *Rev. Palaeobot. Palynol.* **91**, 23–34 (1996).
- Palmer, K. V. W. & Brann, D. C. Catalogue of the Paleocene and Eocene mollusca of the southern and eastern United States: Part I. Pelecypoda, Amphineura, Pteropoda, Scaphopoda, and Cephalopoda. *Bull. Am. Paleontol.* **48**, 1–443 (1965).
- Palmer, K. V. W. & Brann, D. C. Catalogue of the Paleocene and Eocene mollusca of the southern and eastern United States: Part II. Gastropoda. *Bull. Am. Paleontol.* **48**, 471–1027 (1966).
- Dockery, D. T. Lower Oligocene Bivalvia of the Vicksburg Group. *Mississippi Dept Natural Resour. Bur. Geol. Bull.* **123**, 1–261 (1982).
- MacNeil, F. S. & Dockery, D. T. Lower Oligocene Gastropoda, Scaphopoda, and Cephalopoda of the Vicksburg Group in Mississippi. *Mississippi Dept Natural Resour. Bur. Geol. Bull.* **124**, 1–415 (1984).
- Hansen, T. in *Eocene-Oligocene Climatic and Biotic Evolution* (eds Prothero, D. R. & Berggren, W. A.) 341–348 (Princeton Univ. Press, Princeton, New Jersey, 1992).
- Frederiksen, N. O. Sporomorph biostratigraphy, floral changes, and paleoclimatology, Eocene and Earliest Oligocene of the Eastern Gulf Coast. *US Geol. Surv. Prof. Pap.* **1448**, 1–68 (1988).
- Seisser, W. G. Paleogene sea levels and climates: U.S.A. Eastern Gulf Coastal Plain. *Paleogeogr. Paleoclimatol. Paleoecol.* **47**, 261–275 (1984).
- Kalish, J. M. ^{13}C and ^{18}O isotopic disequilibria in fish otoliths: metabolic and kinetic effects. *Mar. Ecol. Prog. Ser.* **75**, 191–203 (1991).
- Kalish, J. M. Oxygen and carbon stable isotopes in the otoliths of wild and laboratory-reared Australian salmon (*Aripis trutta*). *Mar. Biol.* **110**, 37–47 (1991).
- Iacumin, P., Bianucci, G. & Longinelli, A. Oxygen and carbon isotopic composition of fish otoliths. *Mar. Biol.* **113**, 537–542 (1992).
- Patterson, W. P., Smith, G. R. & Lohmann, K. C. in *Climate Change in Continental Isotopic Records* (eds Swart, P., Lohmann, K. C., McKenzie, J. & Savin, S.) 191–202 (AGU Monograph 78, American Geophysical Union, Washington DC, 1993).
- Valentine, J. W. *Evolutionary Paleogeology of the Marine Biosphere* (Prentice-Hall, Englewood Cliffs, New Jersey, 1973).
- Vermeij, G. J. *Biogeography and Adaptation* (Harvard Univ. Press, Cambridge, Massachusetts, 1978).
- Miller, K. G. in *Eocene-Oligocene Climatic and Biotic Evolution* (eds Prothero, D. R. & Berggren, W. A.) 160–177 (Princeton Univ. Press, Princeton, New Jersey, 1992).
- Zachos, J. C., Breza, J. R. & Wise, S. W. Early Oligocene ice-sheet expansion on Antarctica: Stable isotope and sedimentological evidence from Kerguelen Plateau, southern Indian Ocean. *Geology* **20**, 569–573 (1992).
- Zachos, J. C., Lohmann, K. C., Walker, J. C. G. & Wise, S. W. Abrupt climate change and transient climates during the Paleogene: A marine perspective. *J. Geol.* **101**, 191–213 (1993).
- Lear, C. H., Elderfield, H. & Wilson, P. A. Cenozoic deep-sea temperatures and global ice volumes from Mg/Ca in benthic foraminiferal calcite. *Science* **287**, 269–272 (2000).
- Dettman, D. L. & Lohmann, K. C. in *Climate Change in Continental Isotopic Records* (eds Swart, P. K., Lohmann, K. C., McKenzie, J. & Savin, S.) 153–163 (AGU Monograph 78, American Geophysical Union, Washington DC, 1993).
- NOAA/PMEL World Ocean Atlas [online] (cited Nov. 1999) <http://ferret.wrc.noaa.gov/fbin/climate_server> (1994).
- Weidman, C. R., Jones, G. A. & Lohmann, K. C. The long-lived mollusc *Arctica islandica*: A new paleoceanographic tool for the reconstruction of bottom temperatures for the continental shelves of the Northern Atlantic Ocean. *J. Geophys. Res.* **99**, 18305–18314 (1994).
- Grossman, E. L. & Ku, T.-L. Oxygen and carbon isotope fractionation in biogenic aragonite: temperature effects. *Chem. Geol.* **59**, 59–74 (1986).
- Berggren, W. A., Kent, D. V., Swisher, C. C. I. & Aubry, M.-P. A *Revised Cenozoic Geochronology and Chronostratigraphy* 129–212 (Special Publication 54, SEPM, Tulsa, Oklahoma, 1995).
- Baum, J. S., Baum, G. R., Thompson, P. R. & Humphrey, J. D. Stable isotopic evidence for relative and eustatic sea-level changes in Eocene to Oligocene carbonates, Baldwin County, Alabama. *Geol. Soc. Am. Bull.* **106**, 824–839 (1994).
- Yancey, T. E. & Davidoff, A. J. *Paleogene Sequence Stratigraphy of the Brazos River Section, Texas* (Gulf Coast Association of Geological Societies, Austin, Texas, 1994).

Acknowledgements

We thank C. Wurster, L. Wingate and G. Hourigan for analysing samples and assisting with microsampling, and M. Suchter for assistance in microsampling, data entry and analysis; D. Nolf for providing additional Palaeogene otoliths from the Gulf Coast to augment our available samples; and T. Baumiller, B. Wilkinson, R. Buick and P. Wilf for comments on the manuscript. Ideas presented here benefited from discussion with participants of the GSA Penrose Conference on the Eocene/Oligocene boundary held in August 1999. This work was supported by the Michigan Society of Fellows and the NSF.

Correspondence should be addressed to L.C.I. (e-mail: lcivany@syr.edu).

Non-chondritic distribution of the highly siderophile elements in mantle sulphides

Olivier Alard*, William L. Griffin†‡, Jean Pierre Lorand‡, Simon E. Jackson* & Suzanne Y. O'Reilly*

* GEMOC, Department of Earth and Planetary Sciences, Macquarie University, Sydney, New South Wales 2109, Australia

† CSIRO Exploration and Mining, PO Box 136, North Ryde, New South Wales 1670, Australia

‡ Laboratoire de Mineralogie, Muséum National d'Histoire Naturelle de Paris, Unité ESA CNRS no. 7058, 61 Rue Buffon, 75005 Paris, France

The abundances of highly siderophile (iron-loving) elements (HSEs) in the Earth's mantle provide important constraints on models of the Earth's early evolution. It has long been assumed that the relative abundances of HSEs should reflect the composition of chondritic meteorites—which are thought to represent the primordial material from which the Earth was formed. But the non-chondritic abundance ratios recently found in several types of rock derived from the Earth's mantle^{1–3} have been difficult to reconcile with standard models of the Earth's accretion^{4–9}, and have been interpreted as having arisen from the addition to the primitive mantle of either non-chondritic extraterrestrial material or differentiated material from the Earth's core. Here we report *in situ* laser-ablation analyses of sulphides in mantle-derived rocks which show that these sulphides do not have chondritic HSE patterns, but that different generations of sulphide within single samples show extreme variability in the relative abundances of HSEs. Sulphides enclosed in silicate phases have high osmium and iridium abundances but low Pd/Ir ratios, whereas pentlandite-dominated interstitial sulphides show low osmium and iridium abundances and high Pd/Ir ratios. We interpret the silicate-enclosed sulphides as the residues of melting processes and interstitial sulphides as the crystallization products of sulphide-bearing (metasomatic) fluids. We suggest that non-chondritic HSE patterns directly reflect processes occurring in the upper mantle—that is, melting and sulphide addition via metasomatism—and are not evidence for the addition of core material or of 'exotic' meteoritic components.

Until recently, the HSEs (that is, the platinum group elements

(PGEs) + Au + Re) were thought to be present in the Earth's mantle in chondritic relative abundances^{5–8} but at higher levels than expected from models of core formation⁴. This apparent discrepancy is the basis of the late-veneer hypothesis^{4–9}—that is, re-enrichment of the mantle in HSEs by a late meteoritic bombardment, after core formation. However, several recent studies have identified mantle samples with Pd/Ir, Pt/Ir and Rh/Ir ratios higher than chondritic ratios^{1–3,10}. Such non-chondritic ratios have led to speculation that either non-chondritic extraterrestrial material or differentiated outer-core material has been added to the primitive mantle¹. But available data on the ¹⁸⁷Os/¹⁸⁸Os ratios of the upper mantle are generally consistent with chondritic Re/Os ratios⁹, suggesting that the parent/daughter (Re/Os) ratio of the upper mantle has been chondrite-like (or lower) for a long time¹¹ and that supra-chondritic HSE ratios are probably related to 'young' processes. The lack of detailed constraints on HSE behaviour within the lithospheric mantle has resulted in considerable debate on the origin of 'abnormal' mantle material such as the "high-Pd mantle"^{1–3,10}.

Since the study by Mitchell and Keays⁶, sulphides have been regarded as the main host phases for HSEs in the mantle. This assumption has been confirmed by *in situ* proton microprobe analysis^{12,13} and by analyses of separated sulphide fractions^{5,10,14}. These studies reported contrasting primitive-mantle-normalized (PM-normalized) PGE patterns (Figs 1, 2). Sulphide inclusions in diamonds of peridotitic paragenesis are characterized by high I-PGE contents (I-PGE = Ir, Os, Ru and Rh) and sub-chondritic Pd/Ir ratios¹², in contrast to the lower I-PGE contents and supra-chondritic Pd/Ir ratios of sulphides from the Lherz orogenic peridotite¹⁰ (Fig. 2).

We have determined the contents of PGEs and Au in sulphides from several types of mantle-derived rocks, using laser ablation microprobe inductively coupled plasma mass spectrometry (LAM-ICPMS; Table 1). This technique can analyse individual grains of mantle sulphides, with much lower detection limits than the proton microprobe.

The analysed samples represent the main types of sulphide occurrence and the range of sulphur contents in mantle-derived peridotites, from both cratonic and circum-cratonic mantle lithosphere. The cratonic samples investigated here are olivine xenocrysts—with forsterite contents %Fo (= 100Mg/(Mg+Fe)) of 89–93 from the Udachnaya kimberlite (Siberian craton, Russia) and a garnet lherzolite xenolith (FS-8) from the Frank Smith kimberlite (Kapaavaal craton). The sulphides in the olivine xenocrysts are rounded blebs displaying finely interfingered Ni-poor and Ni-rich

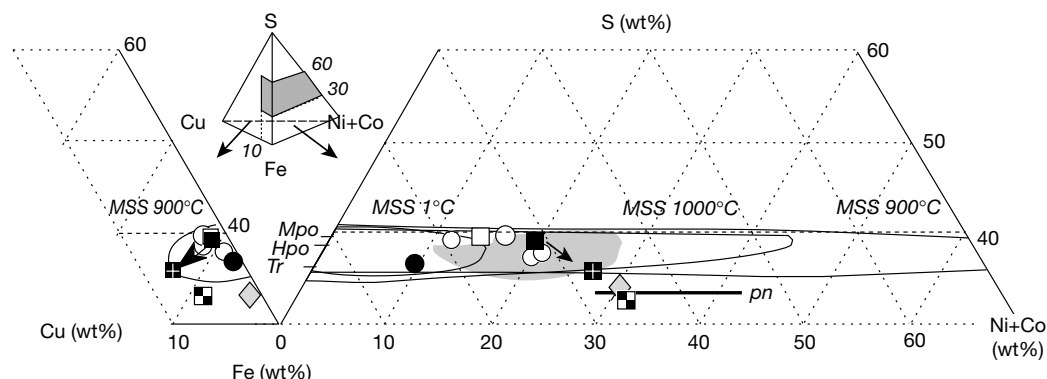


Figure 1 Bulk sulphide compositions plotted in the Fe–Ni–S–Cu system. Open circles, sulphide in olivine macrocrysts from Udachnaya; filled circles, average garnet-enclosed MSS in FS-8; open square, average olivine-enclosed sulphide in S-poor Mt Quincan xenoliths. Filled square and white-crossed filled square are olivine-enclosed and interstitial sulphides, respectively, in Mt Gambier xenoliths; chequerboard symbol,

average interstitial sulphides in AR10 xenolith; grey diamond, interstitial sulphide in 84-1 lherzolite from the Lherz orogenic massif. High-temperature (1,100–900 °C) phase relations are after ref. 28. Grey field, bulk enclosed sulphides in mantle xenoliths^{15–17}. End-member phases: Tr, troilite; Mpo, monoclinic pyrrhotite; Hpo, hexagonal pyrrhotite; pn, pentlandite (compositional range after ref. 19).

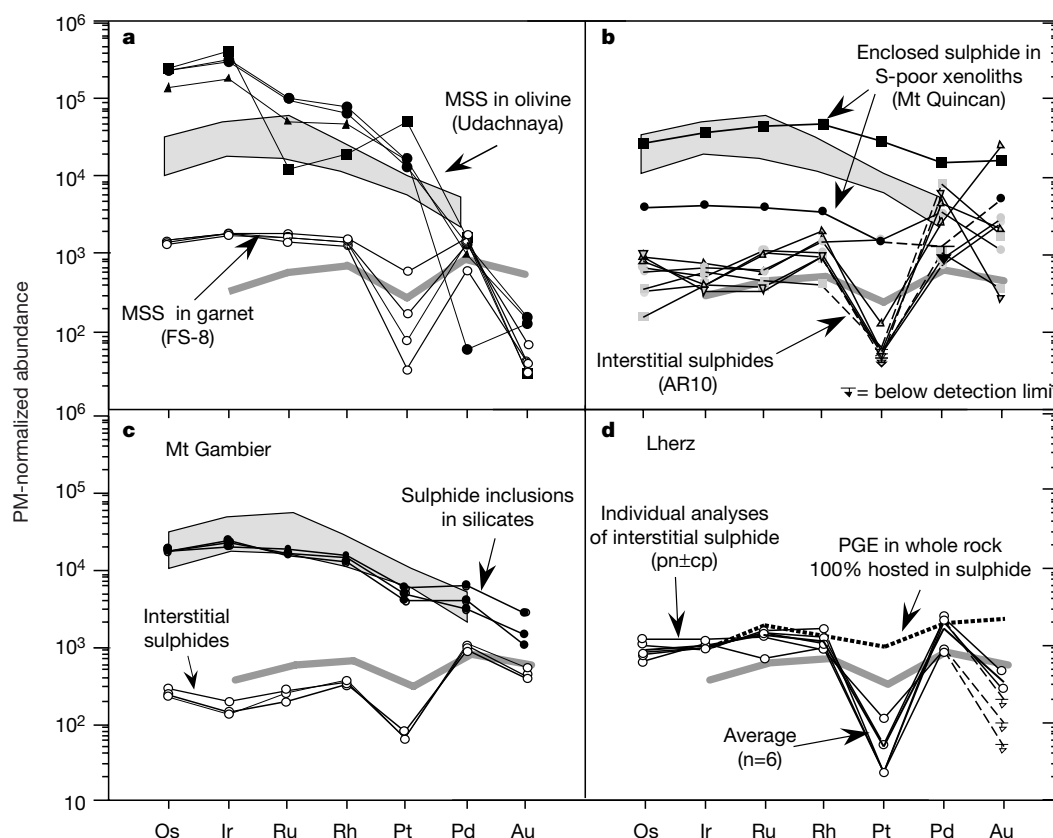


Figure 2 Primitive-mantle²⁹-normalized HSE abundances of sulphide. Grey area, MSS enclosed in peridotitic diamond¹³; thick grey line, bulk sulphide fraction from Lherz⁹. **a**, MSS in cratonic samples; filled symbols are olivine-enclosed MSS from Udachnaya (circles, Ol61; squares, Ol62; triangles, Ol60); open circles are individual garnet-enclosed sulphides from sample FS-8. **b**, Sulphide in xenoliths from basalts; filled symbols are olivine-enclosed MSS in harzburgite ($\text{Al}_2\text{O}_3 = 0.33\%$) A4/37 (squares) and in a lherzolite

($\text{Al}_2\text{O}_3 = 3.64\%$) xenolith A4/39 (circles). **c**, Silicate-enclosed MSS (filled circles), and interstitial Cu-rich Pn (open circles) from the same xenolith (Mt Gambier, sample Gam VL9, $\text{Al}_2\text{O}_3 = 1.85\%$). **d**, Interstitial pn in 84-1 lherzolite (Lherz massif); thick dashed line assumes all PGE in 84-1 (whole rock²; WR) resides in sulphide having S content $\sim 33.5 \text{ wt}\%$ (S_{WR} , 232 p.p.m.); cp, chalcopyrite.

monosulphide solid solution (MSS) and pentlandite, with minor chalcopyrite at the rim of the blebs. They have an MSS bulk composition encompassing the various fields described for enclosed MSS in xenoliths^{15–17} (Fig. 1). FS-8 is a high-temperature ($\sim 1,300^\circ\text{C}$), very fertile (high-Ti, low-Cr garnet, $\% \text{Fo} \approx 89$), sheared garnet lherzolite. Sulphides are included in garnet as large sub-circular blebs (up to $300 \mu\text{m}$ across) showing structures like those of the Udachnaya sulphides but with a higher proportion of Fe-hydroxide alteration (Fig. 1).

MSS enclosed in olivine macrocrysts and garnet has high contents of the more refractory I-PGE group and decreasing PM-normalized contents of the more volatile PGEs (that is, Pt and Pd: P-PGE). They show arch-shaped PM-normalized HSE patterns similar to those of MSS in peridotitic diamond¹² (Fig. 2a). One MSS bleb in olivine has Os and Ir contents as high as 878 and 1,374 p.p.m. ($>100,000 \times \text{PM}$), respectively, and a strong Pd-depletion ($(\text{Pd}/\text{Ir})_{\text{PM}} \approx 0.01$). Os and Ir contents are correlated with $\% \text{Fo}$ of the enclosing olivine; Pd/Ir is negatively correlated with $\% \text{Fo}$. The more fertile peridotite FS-8 displays a lower Os-Ir content and a less Pd-depleted pattern.

Sulphides in mantle-peridotite xenoliths entrained by alkali basalts have been studied using samples from three eastern Australian volcanic provinces (Mt Quincan, Allyn River, Mt Gambier). Xenoliths from Mt Quincan (Northern Queensland) are S-poor ($S \leq 27 \text{ p.p.m.}$) irrespective of their fertility indices ($0.7 < \text{Al}_2\text{O}_3\% < 4.1$). Rare sulphides occur as rounded blebs of MSS ($< 50 \mu\text{m}$) enclosed in olivine. The composition and mineralogy of these blebs

are similar to those in the Udachnaya olivine macrocrysts (Fig. 1). They also display PM-normalized HSE patterns similar to those of MSS inclusions (Fig. 2b) in the 'cratonic' xenoliths and diamonds: high I-PGE contents ($100 > \text{Os} > 15 \text{ p.p.m.}$) and low P-PGE contents ($(\text{Pd}/\text{Ir})_{\text{PM}} < 0.5$).

Xenoliths from Allyn River (New South Wales) are strongly enriched in light rare-earth elements (LREEs; $(\text{La}/\text{Sm})_{\text{PM}} > 10$) and have high S contents—up to 330 p.p.m. AR-10, the most S-rich of this suite, displays large anhedral sulphide grains (up to $350 \mu\text{m}$) that enclose euhedral secondary phases (olivine, clinopyroxene, spinel) in interstitial clusters formed by the introduction of fluids. The sulphides are mainly pentlandite $\pm$ pyrrhotite with exsolved lamellae of pentlandite, and common chalcopyrite. The bulk chemistry is a Cu-rich pentlandite (Fig. 1). Such sulphides are common to all the low-temperature ($< 1,050^\circ\text{C}$) LREE- and S-enriched xenoliths from this locality. HSE patterns are characterized by low I-PGE contents ($0.5 < \text{Os} < 3.5 \text{ p.p.m.}$) but variable enrichment in Pd and Au relative to the I-PGE ($(\text{Pd}/\text{Ir})_{\text{PM}} = 1\text{--}20$; Fig. 2b).

Xenoliths from Mt Gambier (South Australia) have S contents ranging up to 130 p.p.m. and contain two types of sulphide. Rounded MSS enclosed in olivine are similar to those in the S-poor xenoliths from Mt Quincan. Interstitial sulphides are larger than the olivine-hosted inclusions ($70\text{--}250 \mu\text{m}$) and characterized by high proportions of chalcopyrite and pentlandite. The bulk composition is Cu-rich pentlandite as in AR-10 (Fig. 1). The silicate-enclosed sulphides display the arch-shaped PM-normalized

patterns typical of other silicate-enclosed MSS (Fig. 2c). In addition to their similarities in major elements, the interstitial sulphides in the Mt Gambier xenoliths display the same HSE patterns as the metasomatic sulphides in AR-10 (Fig. 2c): high Pd/Ir ratios, negative Pt anomalies relative to Rh and Pd, and low I-PGE contents ($200\text{--}300 \times \text{PM}$). These similarities suggest that the interstitial sulphides in Mt Gambier xenoliths are also of metasomatic origin. Many xenoliths from alkali basalts (Eastern Australia and Massif Central, France) contain these two types of sulphides.

A spinel lherzolite (84-1; $\text{Al}_2\text{O}_3\% = 3.64$) from the Lherz orogenic massif (Pyrenées, France) is S-rich ($S = 232 \text{ p.p.m.}$) and contains only the interstitial pentlandite grains commonly described in orogenic peridotites¹⁸. Chalcopyrite and pyrrhotite are minor phases. The HSE patterns of the pentlandite are also similar to those of AR-10 (Fig. 2d), with weakly fractionated I-PGE patterns (about $1,000 \times \text{PM}$) and negative Pt anomalies, but less Pd enrichment ($0.7 < (\text{Pd}/\text{Ir})_{\text{PM}} < 2.3$). LAM-ICPMS analyses show patterns similar to the analysis of the bulk separated sulphide¹⁰. However, our analyses show deeper negative Pt anomalies ($(\text{Pt}/\text{Ir})_{\text{PM}} < 0.05$) and much lower Au contents.

The mass balance between whole-rock PGE abundances² and average sulphide content in Lherz peridotites suggests that Os, Ir, Ru, Rh and Pd can be fully accounted for by the sulphide (Fig. 2d). However, this is not the case for Au and Pt. The discrepancy between the bulk analysis and the *in situ* analysis suggests that Au and Pt either enter other minor sulphide phases (for example, chalcopyrite) too small to be analysed by LAM-ICPMS, or occur as discrete minerals. During the ablation runs, peaks of Pt appeared as we drilled through to the edge of the sulphide grains, suggesting that discrete Pt phases occur at the sulphide–silicate boundaries. This nugget effect is observed both in MSS showing alteration features and in pentlandite-dominated interstitial sulphides. The Pt nugget effect may be the result of secondary processes such as serpentinization, which dramatically reduces the oxygen fugacity¹⁹; under such conditions Pt will be more stable as an alloy than in solid solution in base metal sulphide. Experimental studies show that pentlandite can accommodate levels of Ru, Rh and Pd of a few per cent, but only trace amounts of Pt (refs 18, 20), so Pt alloys such as Pt_3Fe may exsolve from pentlandite at low temperatures.

These data confirm that HSE are strongly concentrated in mantle

sulphides and not in the coexisting silicates. The time-resolved signals observed during laser ablation of mantle sulphides show few irregularities. This strongly suggests that the PGEs, with the common exception of Pt, occur in the lattice of the sulphide phase, rather than as irregularly disseminated clusters as reported for crustal sulphides²¹.

In summary, two texturally and compositionally distinct types of sulphides occur in mantle-derived peridotites: (1) rounded MSS enclosed in silicates, usually in depleted, S-poor samples, and (2) interstitial sulphides, often rich in Ni or Cu. Both types commonly occur in a single sample. The complementary nature of their HSE patterns suggests that these two sulphide types are related through a single process: incongruent partial melting of sulphide having a 'primitive' chondritic HSE pattern. In this model the MSS represent residues^{15–17}, while the interstitial/metamorphic sulphides could represent partial melts^{3,22}. These strong differences between MSS and interstitial S-poor sulphides agree with experimental data²³ that suggest a strong affinity of the I-PGE for the MSS structure, while Pt and Pd partition into the coexisting sulphide liquids. The MSS HSE patterns also suggest that the PGE behaviour can be better explained by MSS–liquid partitioning. This process can account for the incompatible behaviour of Pd—as observed here in sulphides, or in S-depleted peridotites (that is, orogenic harzburgites²)—which cannot be reconciled with experimentally determined sulphide–liquid/silicate–liquid partition coefficients^{24,25}.

Whatever the process, it is most unlikely that the two types of sulphide described here represent addition either of core material, or of distinct meteorite populations, to the upper mantle. The movement of the metasomatic sulphides analysed here has the potential to produce supra-chondritic Pd/Ir ratios in the lithospheric mantle. Therefore we suggest that the non-chondritic relative abundances of HSE recently found in various mantle samples are due to sulphide differentiation within the mantle, by partial melting and sulphide addition through melt/rock reaction, and do not preclude a chondritic HSE abundance for the Earth.

On the other hand, we have found no mantle-derived sulphides with chondritic HSE patterns. This suggests that reported whole-rock samples with near-chondritic relative abundances of the HSEs may represent mixtures of the two types of sulphide described here. Also, preliminary data^{26,27} show that these two generations of

Table 1 HSE contents of sulphide

Samples	Sulphide type	Os (p.p.m.)	Ir (p.p.m.)	Ru (p.p.m.)	Rh (p.p.m.)	Pt (p.p.m.)	Pd (p.p.m.)	Au (p.p.m.)
Siberia; olivine macrocrysts, Udachnaya pipe								
Ol 61 a1*	e-ol	824 ± 81	1,049 ± 60	523 ± 29	58 ± 4	95 ± 5	5.6 ± 0.3	0.16 ± 0.01
	(MSS + Cp)							
Ol 61 a2*	e-ol(MSS)	835 ± 82	1,104 ± 80	531 ± 31	73 ± 5	118 ± 6	0.23 ± 0.04	0.13 ± 0.01
Kapaavaal; high-temperature sheared fertile garnet-peridotite, Frank Smith pipe (Gt: $\text{TiO}_2 = 1.28$, $\text{Cr}_2\text{O}_3 = 1.8 \text{ wt}\%$)								
FS-8 a1	e-gt (altered MSS)	4.8 ± 0.4	5.5 ± 0.2	8.4 ± 0.5	1.3 ± 0.1	1.2 ± 0.06	5.6 ± 0.4	0.07 ± 0.01
FS-8 b1	e-gt (altered MSS)	5.3 ± 0.4	5.8 ± 0.2	8.2 ± 0.4	1.3 ± 0.1	0.56 ± 0.03	5.9 ± 0.4	0.03 ± 0.01
Eastern Australia; Mantle peridotite xenoliths in alkali basalts (AR: Allyn River, Gam: Mt Gambier, 94A4: Mt Quincan)								
94 A4-37	e-ol (MSS)	15 ± 1	15 ± 1	22 ± 1	3.5 ± 0.3	11 ± 1	<5.3	5.7 ± 0.4
94 A4-39	e-ol (MSS)	100 ± 7	131 ± 8	258 ± 16	4.8 ± 0.4	231 ± 9	65 ± 6	18 ± 1
AR-10a	i (Cu-rich pn)	1.3 ± 0.2	1.9 ± 0.2	2.3 ± 0.1	0.39 ± 0.06	<0.46	35 ± 2	1.8 ± 0.1
AR-10f	i (Cu-rich pn)	2.1 ± 0.2	2.4 ± 0.2	3.2 ± 0.2	1.4 ± 0.2	12 ± 1	14 ± 1	1.19 ± 0.07
GamVL9 b	e-ol (MSS)	62 ± 6	76 ± 5	75 ± 6	11 ± 1	28 ± 3	15 ± 1	1.0 ± 0.1
GamVL9 a	i (Cu-rich pn)	0.76 ± 0.09	0.40 ± 0.05	1.2 ± 0.1	0.35 ± 0.07	0.45 ± 0.05	4.3 ± 0.4	0.56 ± 0.05
Pyrenées; lherzolite 84-1 ($\text{Al}_2\text{O}_3 = 3.4 \text{ wt}\%$, $S = 232 \text{ p.p.m.}$), Lherz orogenic massif								
84-1	i (pn)	3.0 ± 0.7	3.2 ± 0.4	6.4 ± 1.7	1.03 ± 0.3	0.34 ± 0.29	6.4 ± 2.7	<0.05 ± 0.45
	average of 6							
PGE-A standard (NIS beads)								
Average of 44		192 ± 12	215 ± 10	208 ± 9	226 ± 10	124 ± 7	277 ± 13	219 ± 11
Typical detection limit		0.02	0.02	0.07	0.03	0.04	0.04	0.04

All values are shown $\pm 1\sigma$. Abbreviations: i, interstitial; e, enclosed; ol, olivine; gt, garnet; MSS, mono-sulphide solid solution; pn, pentlandite; Cp, chalcopyrite. The contents of six PGEs and Au were determined using an in-house Nd:YAG laser-ablation system linked to a Perkin Elmer Sciex ELAN 6000 ICP-MS. A quenched NIS bead doped with PGE and other chalcophile elements was used as an external standard. This standard (PGE-A) matches the sulphide matrix, allowing straightforward data processing. The average $\pm 1\sigma$ of 44 analyses of PGE-A analysed as an unknown is given along with typical detection limits (using He as carrier gas). Typical analytical conditions: 40–60 μm spot diameter, 4 Hz laser frequency, and a beam energy around 0.5 mJ. Sulphide in olivine 61 has been analysed twice by laser ablation ICP-MS (a1, a2) and by proton microprobe (C. Ryan, personal communication); Os 892 $\pm$ 126 p.p.m., Ir = 712 $\pm$ 70 p.p.m., Ru = 517 $\pm$ 19 p.p.m., Rh = 72 $\pm$ 8 p.p.m., Pt < 55 p.p.m., Pd < 12 p.p.m. Further details of these analyses are available; see Supplementary Information.

sulphides can have different Os isotope compositions. Care must therefore be exercised in interpreting whole-rock PGE and Re/Os data without detailed information on these sulphides. □

Received 29 February; accepted 23 August 2000.

1. Snow, J. E. & Schmidt, G. Constraints on Earth accretion deduced from noble metals in the oceanic mantle. *Nature* **391**, 166–169 (1998).
2. Lorand, J. P., Gros, M. & Pattou, L. Fractionation of platinum-group elements in the upper mantle: a detailed study in Pyrenean orogenic peridotites. *J. Petrol.* **40**, 951–987 (1999).
3. Rehkämer, M. *et al.* Non-chondritic platinum-group element ratios in oceanic mantle lithosphere: Petrogenetic signature of melt percolation. *Earth Planet. Sci. Lett.* **172**, 65–81 (1999).
4. Borisov, A., Palme, H. & Spettel, B. Solubility of palladium in silicate melts: implications for core formation in the Earth. *Geochim. Cosmochim. Acta* **58**, 705–716 (1994).
5. Jagoutz, E. H. *et al.* The abundance of major, minor and trace elements in the earth's mantle as derived from primitive ultramafic nodules. *Proc. Lunar. Planet. Sci. Conf. X*, 2031–2050 (1979).
6. Mitchell, R. H. & Keays, R. R. Abundance and distribution of gold, palladium and iridium in some spinel and garnet lherzolites: implications for the nature and origin of precious metal-rich intergranular components in the upper mantle. *Geochim. Cosmochim. Acta* **45**, 2425–2445 (1981).
7. Morgan, J. W. Ultramafic xenoliths: clues to earth's late accretionary history. *J. Geophys. Res.* **91**, 12375–12387 (1986).
8. O'Neill, H. S. C. The origin and the early history of the Earth—A chemical model. Part 2: The Earth. *Geochim. Cosmochim. Acta* **55**, 1159–1172 (1991).
9. Meisel, T., Walker, R. J. & Morgan, J. W. The osmium isotopic composition of the Earth's primitive upper mantle. *Nature* **383**, 517–520 (1996).
10. Pattou, L., Lorand, J. P. & Gros, M. Non-chondritic platinum-group element ratios in the Earth's mantle. *Nature* **379**, 712–715 (1996).
11. Shirey, S. B. & Walker, R. J. The Re-Os isotope system in cosmochemistry and high-temperature geochemistry. *Annu. Rev. Earth. Planet. Sci.* **26**, 423–500 (1998).
12. Bulanova, G. P., Griffin, W. L., Ryan, C. G., Shestakova, O. Ye. & Barnes, S. J. Trace element in sulfide inclusions from Yakutian diamonds. *Contrib. Mineral. Petrol.* **124**, 111–125 (1996).
13. Guo, J., Griffin, W. L. & O'Reilly, S. Y. Geochemistry and origin of sulfide minerals, in mantle xenoliths: Qilin, southeastern China. *J. Petrol.* **40**, 1125–1149 (1999).
14. Hart, S. R. & Ravizza, G. E. in *Reading the Isotopic Code* (eds Basu, A. & Hart, S. R.) 123–134 (American Geophysical Union, Washington DC, 1996).
15. Lorand, J. P. & Conqu  r, F. Contribution a l'  tude des sulfures dans les enclaves de lherzolites a spinelle des basaltes alcalins (Massif Central et Languedoc, France). *Bull. Min  ral.* **106**, 585–606 (1983).
16. Dromgoole, E. L. & Pasteris, J. D. in *Mantle Metasomatism and Alkaline Magmatism* (eds Morris, E. & Pasteris, J. D.) 25–46 (Special Paper 215, Geological Society of America, Washington DC, 1987).
17. Szab  , C. S. & Bodnar, R. J. Chemistry and origin of mantle sulfides in spinel peridotite xenoliths from alkaline basaltic lavas, Nograd-Gomor Volcanic Field, northern Hungary and southern Slovakia. *Geochim. Cosmochim. Acta* **59**, 3917–3927 (1995).
18. Mackovicky, M., Mackovicky, E. & Rose-Hansen, J. in *Metallogeny of Basic and Ultrabasic Rocks* (eds Gallagher, M. J., Ixer, R. A., Neary, C. R. & Prichard, H. M.) 415–425 (Inst. Min. Metl., London, 1986).
19. Lorand, J. P. Are spinel lherzolite xenoliths representative of the abundance of sulfur in the upper mantle? *Geochim. Cosmochim. Acta* **54**, 1487–1492 (1989).
20. Ballhaus, C. & Ryan, C. G. Platinum-group elements in the Merensky Reef. 1. PGE in solid solution in base metal sulfides and the down-temperature equilibration history of Merensky ores. *Contrib. Mineral. Petrol.* **122**, 241–251 (1995).
21. Balhaus, C. & Sylvester, P. Noble metal enrichment processes in the Merensky Reef, Bushveld complex. *J. Petrol.* **41**, 546–561 (1999).
22. Luguet, A. & Lorand, J. P. Min  ralogie des sulfures de Fe-Ni-Cu dans les p  ridotites abyssales de la zone Mark (ride m  dio-Atlantique, 20–24  N). *C.R. Acad. Sci. Paris* **329**, 637–644 (1999).
23. Li, C., Barnes, S. J., Mackovicky, E., Rose-Hansen, J. & Mackovicky, M. Partitioning of nickel, copper, iridium, rhodium, platinum and palladium between monosulfide solution and sulphide liquid: Effects of composition and temperature. *Geochim. Cosmochim. Acta* **60**, 1231–1238 (1996).
24. Peach, C. L., Mathez, E. A., Keays, R. R. & Reeves, S. J. Experimentally-determined sulfide melt-silicate melt partition coefficients for iridium and palladium. *Chem. Geol.* **117**, 361–377 (1994).
25. Fleet, M. E., Crockett, J. H., Liu, M. & Stone, W. E. Laboratory partitioning of platinum-group elements (PGE) and gold with application to magmatic sulfide–PGE deposits. *Lithos* **47**, 127–142 (1999).
26. Burton, K. W., Chiano, P., Birch, J.-L. & Allegre, C. J. Osmium isotope disequilibrium between mantle minerals in a spinel-lherzolite. *Earth Planet. Sci. Lett.* **172**, 311–322 (1999).
27. Alard, O., Pearson, N. J., Griffin, W. L., Graham, S. & Jackson, S. E. in *Beyond 2000, New Frontiers in Isotope Geoscience, Extended Abstract Volume 1–5* (Lorne, Australia, 2000).
28. K  llerud, G., Yund, R. A. & Moh, G. H. Phase relation in the Cu-Fe-Ni, Cu-Ni-S and Fe-Ni-S systems. *Econ. Geol.* **4**, 323–343 (1969).
29. McDonough, W. F. & Sun, S. S. The chemical composition of the Earth. *Chem. Geol.* **120**, 223–253 (1995).

Supplementary information is available on *Nature's* World-Wide Web site (<http://www.nature.com>) or as paper copy from the London editorial office of *Nature*.

Acknowledgements

We thank F.R. (Joe) Boyd, S. Talnikova, Y. Barashkov and W.J. Powell for providing samples, and N.J. Pearson, C. Lawson and A. Sharma for assistance with the analytical facilities. We thank Y. Lahaye for comments on an earlier version of this Letter, and R. Carlson and M. Reh  k  mper for comments on the final version. This is a GEMOC National Key Centre publication.

Correspondence and requests for materials should be addressed to O.A. (e-mail: oalard@laurel.ocs.mq.edu.au).

Starch grains reveal early root crop horticulture in the Panamanian tropical forest

Dolores R. Piperno*, Anthony J. Ranere†, Irene Holst* & Patricia Hansell†

* Smithsonian Tropical Research Institute, Box 2072, Ancon, Balboa, Republic of Panama

† Department of Anthropology, Temple University, Philadelphia, Pennsylvania 19122, USA

Native American populations are known to have cultivated a large number of plants and domesticated them for their starch-rich underground organs¹. Suggestions^{2,3} that the likely source of many of these crops, the tropical forest, was an early and influential centre of plant husbandry have long been controversial^{4–6} because the organic remains of roots and tubers are poorly preserved in archaeological sediments from the humid tropics. Here we report the occurrence of starch grains identifiable as manioc (*Manihot esculenta* Crantz), yams (*Dioscorea* sp.) and arrowroot (*Maranta arundinacea* L.) on assemblages of plant milling stones from preceramic horizons at the Aguadulce Shelter, Panama, dated between 7,000 and 5,000 years before present (BP). The artefacts also contain maize starch (*Zea mays* L.), indicating that early horticultural systems in this region were mixtures of root and seed crops. The data provide the earliest direct evidence for root crop cultivation in the Americas, and support an ancient and independent emergence of plant domestication in the lowland Neotropical forest.

The site (Ag-13) is located on the Pacific coastal plain of Central Panama approximately 17 km from the sea⁷. Today the region receives 1,600 mm of precipitation annually, distributed on a highly seasonal basis, and has a mean annual temperature of 26  C. The potential vegetation is a deciduous tropical forest. Excavations carried out in 1973–75 and 1997 uncovered the remains of three distinct human occupations. Zone D, the oldest, is a yellow (inside the dripline) or red (beyond the dripline) silty clay on top of weathered bedrock. It has uncalibrated ¹⁴C dates of 10,725 ± 80 BP (NZA-10930), 10,529 ± 184 BP (NZA-10930), 8,423 ± 79 BP

Table 1 Number and types of starch grains on milling stones from the Aguadulce Shelter

Catalogue number	Manioc	Maize	<i>Dioscorea</i> spp.	Arrowroot	Other	Total <i>n</i>
Zone C, bottom						
Block 3 (42)	8	1	3	0	2	14
Block 2 (26b)	1	2	0	0	0	3
Block 3 (38)	0	1	0	0	0	1
1S3E (439)	0	7	0	0	0	0
Zone C, middle						
Block 2 (26a)	5	15	2	2	5	29
Zone C, top						
1S4E (350)	0	25	16	0	13	54
Zone B, bottom						
Block 2 (23)	0	4	0	0	0	4
Zone B, middle						
1S6E (519)	0	3	0	0	9	12
Zone B, top						
3W1N (243)	0	7	0	0	0	7
ONOE (402)	0	7	0	0	0	7
1S5E (316–1)	0	3	0	0	0	3
3W7N (318)	0	2	0	0	1	3

Block/geographic orientation indicates the excavation unit. Field catalogue number for each artefact is given in parentheses. All artefacts except tool 26a, a boulder milling stone base, are edge-ground cobbles. Artefacts with a block number are from the 1973–1975 excavations, and those with a geographic orientation are from the 1997 excavations. The single bell-shaped grain from tool 26b is very likely to be from manioc.

(NZA-9625), $7,061 \pm 81$ BP (NZA-9624) and $5,560 \pm 80$ BP (UCR-3462), all direct accelerator mass spectrometry (AMS) dates on sediment phytoliths^{8,9}. The $5,560 \pm 80$ BP date was run on a very small sample that was recovered in a soil column directly underneath a larger sample dating to $6,910 \pm 60$ BP. The $5,560 \pm 80$ BP date is therefore rejected and Zone D is considered to date from about 11,000 to 7,000 BP because it held a chipped stone industry characterized by bifacial reduction that is present in Panama only during the Paleoindian (about 11,000 to 10,000 BP) and Early Preceramic (about 10,000 to 7,000 BP) periods¹⁰. No milling stones were present.

Zone C, a dark brown clayey silt with angular rock, has two AMS ¹⁴C phytolith dates of $6,910 \pm 60$ BP (UCR-3417) (cited above), at its base, and $6,207 \pm 60$ BP (NZA-10929) stratigraphically higher. A terminal date is estimated to be 5,000 BP. The initiation of this zone marks a great increase in the intensity of human occupation and the rate of sediment accumulation. As in other sites from the region dated to this period, bifacial reduction disappears from the chipped-stone industry, which is now characterized by expedient core reduction and bipolar reduction¹⁰. Milling stones, called edge-ground cobbles because the grinding facet is located on the narrow edge of the stone, first occur near the base of this zone, together with boulder milling stone bases. On the basis of use wear and replicative studies, this distinctive tropical forest tool complex may have been used to process roots and tubers^{11,12}. Faunal remains, including animal bone and shell, are encountered for the first time in Zone C and are abundant and well preserved in sediments underneath the overhang. Ceramics are absent.

Zone B, a tan silt with much less angular rock than Zone C has a single AMS phytolith date of $4,250 \pm 60$ BP (UCR-3418), and is marked by the introduction of 'Monagrillo' pottery, the earliest found in Central America¹³. The zone is dated from about 5,000 to 3,000 BP, on the basis of ¹⁴C dates from the type-site, the Monagrillo shell midden, and others containing this ceramic tradition in Panama¹³. Milling stones from this zone are similar to those recovered from Zone C, but tend to be more heavily worn.

Preliminary analysis of milling stones from the first set of excavations in 1973–75 indicated that they contained starch grains from a variety of plants¹⁴. However, the sample size was small, we lacked sediments that had encased the tools in the ground to confirm that the tools were a primary source of the grains, and we

had a limited selection of modern comparative starches to identify domesticated plants. The 1997 excavations were designed to increase the sample of milling stones and starch and to study the context and chronology of archaeological starch grains in more detail by recovering tools in a pristine condition, together with sediments that were in immediate proximity to the artefacts. We developed a protocol for the collection and analysis of archaeological starch grains, and a key for starch grain identification based on a considerably enlarged modern reference collection (see Methods). Our descriptions and measurements of grains from domesticated and wild taxa correspond to those reported in other large studies^{15–19} and indicate that manioc, yams, arrowroot and maize can be distinguished.

A total of 18 edge-ground cobbles and one boulder milling stone base were analysed. Of these, 12 contained starch residue (Table 1). We recovered 13 bell-shaped starch grains with combinations of attributes that only occur in manioc roots from an edge ground cobble (tool 42) and the boulder milling stone base (tool 26a). The starch grains had rounded and unbanded pressure facet edges, centric and open hila, and no demonstrable lamellae. They often had crossed, winged to y-shaped, or stellated fissures. One of these grains, with pressure facets located on the distal end in the form of a corona with five to eight indentations and a stellated fissure, is a highly idiosyncratic form unique to manioc (Fig. 1a). It occurred on tool 42, which was found near the base of Zone C and should date to nearly 7,000 BP. Eight of the other bell-shaped grains on tool 42 and 26a had crossed, winged to y-shaped, or stellated fissures. Such combinations of fissures on bell-shaped grains with centric hila and without lamellae are unknown in species other than manioc¹⁵. The remaining four bell-shaped grains had no or simple fissures. These also occur in lower frequencies than do the other bell-shaped grains in modern manioc, demonstrating that grain type distributions in the archaeological samples correspond to modern distributions. The identification of a domesticated *Manihot* is further supported by the fact that starch grains in its closest living wild relative *Manihot esculenta* spp. *flabellifolia* (Pohl) Ciferri, which is native to Brazil²⁰, exhibit substantial morphological differences (much narrower pressure facets, simple fissures and mostly eccentric hila), and smaller sizes (Fig. 1b) (Table 2).

Ovoid starch grains with a cuneiform-shaped depression,

Table 2 Size of starch grains in domesticated plants and close wild relatives compared with archaeological grains

Modern plants	Taxonomic status	Range and s.d. of mean length (μm)	Range (μm)	n
<i>Manihot esculenta</i> ssp. <i>flabellifolia</i>	Wild ancestor of <i>M. esculenta</i>	8 ± 2.1	4–12	50
<i>Manihot esculenta</i>	Domesticated	13 ± 2.6 to 16 ± 3.4	6–28	600
Panamanian Poaceae (small-grained, 30 species)	Wild	3 ± 0.9 to 11 ± 2.7	2–18	800
Panamanian Poaceae (large-grained, 5 species)	Wild	14 ± 2.9 to 20 ± 4.8	8–30	250
<i>Zea mays</i> spp. <i>parviglumis</i> *	Wild ancestor of <i>Z. mays</i>	8 ± 2.1 to 9 ± 2.1	4–16	200
<i>Zea mays</i>	Domesticated	8 ± 1.5 to 16 ± 4.9	4–24	750
<i>Dioscorea</i> spp.	Wild	17 ± 4.9 to 63 ± 20	8–100	550
<i>Dioscorea trifida</i>	Domesticated	37 ± 8.3 to 48 ± 12.8	24–84	100
Archaeological grains	Tool number	Mean length	Range	n
<i>Manihot esculenta</i> (all Layer C)	42	14	10–24	8
	26a	18	16–20	5
	26b	18	–	1
<i>Manihot esculenta</i>	Layer C combined	17 ± 3.7	10–24	14
<i>Zea mays</i> , Layer C	42	18	–	1
	26a	15	10–24	15
	26b	13	12–14	2
	439	17	10–22	7
	350	18	10–24	25
	38	14	–	1
<i>Zea mays</i>	Layer C combined	16 ± 4.1	10–24	51
<i>Zea mays</i>	Layer B combined	15 ± 4.9	10–22	26
<i>Dioscorea</i> spp. (all Layer C)	42	15	14–16	3
	26	42	24–60	2
	350	36 ± 7.3	30–52	16

Range of mean length of starch grains in modern plants is the range of means in the populations/races of each species studied. s.d. is the standard deviation for grains in the populations/species having the smallest and largest grains. Range is the range for all starch grains sized in these species. The five wild Panamanian Poaceae (large-grained) with starch grains that demonstrate size overlap with maize (*Cenchrus*, *Antherophora*, *Orthocladia*, *Panicum* spp.) have morphologies distinct from maize.

* Four different populations of *Zea mays* spp. *parviglumis* were studied.

eccentric hila, and lamellae diagnostic of tubers from *Dioscorea* spp. were recovered from these two preceramic tools (42 and 26a), and from another (tool 350) retrieved from the top of Zone C that is probably younger in age than the other two (Fig. 1c). Grains from tools 42 and 26a are morphologically like those of modern wild *Dioscorea* species from Panama, whereas the straight distal edges, morphological uniformity and size of the grains from tool 350 make them identical to those of *D. trifida*, the American domesticated yam. An unidentified, free-growing species of yam from Panama possesses very similar grains, precluding a positive identification of *D. trifida*. Differences in the *Dioscorea* starch residues indicate, however, that at least two species of yams were exploited during the preceramic period. The final tuber starch identified on preceramic artefacts is that of arrowroot (*Maranta arundinaceae*), today a minor root crop of the tropical forest, which possesses diagnostic ovoid grains with eccentric hila and proximal fissures.

Starch grains from grass kernels identified as maize, and which in modern maize are characteristic of races with harder endosperms (such as pop/flint corns), were also found on tools from deep levels in Zone C. As with the manioc grains, the starch assemblage morphologic profiles, presence of individual grains diagnostic of maize (Fig. 1d), and grain size all indicate maize, not wild Poaceae. The same types of maize starches present in Zone C occur in Zone B, which may indicate the persistence of varieties with non-floury endosperms in the ceramic-phase subsistence economy at Ag-13.

The sequence of steps used to analyse the tools and sediments produced the following outcomes (see Methods). Step 1 resulted in starch recovery only from the grinding and/or pounding facets of the tools. In the case of the tools from the 1997 excavations where closely associated sediment was analysed, every starch grain except one was recovered in Steps 1–3, showing that sediments sampled from the immediate proximity of the tools but that were not firmly stuck to artefacts, were nearly devoid of starch. Step 5 yielded only a single starch grain, further indicating that starch preservation in the sediments is very poor, and no starch was recovered from the analysis of four non-artefactual stones and associated sediments.

It appears that starch grains at Ag-13 had a higher probability of survival when they were lodged inside small crevices and cracks on the surfaces of artefacts and, thus, derived protection from decay. Starch research in these contexts opens up a new avenue of palaeoethnobotanical investigation in the Neotropical forest. Formerly, root crop cultivation was indirectly inferred by the presence of edge-ground cobbles and other artefacts that were

thought, on the basis of replicative studies of tool use or an ethnographic analogy, to be used for tuber processing^{11,12}. We cannot rule out the possibility that some of the starch grains on the artefacts represent secondary deposition from the site's sediments after the tools were discarded. Nevertheless, the patterns demonstrated by the starch frequencies on tools and in sediments suggest that the tools were used to process the plants identified. Comparing starch in soils and tools has been used in the Old World as a measure of artefact use²¹.

The starch remains from tuber crops and maize support phytolith and pollen evidence from the Aguadulce Shelter and lake cores in Panama for the presence of well-developed agricultural systems using a variety of crop plants by 7,000–6,000 BP^{22,23}. The Aguadulce starch grains represent the earliest recorded occurrence of manioc in the Americas. Manioc, a staple food in the tropics, was previously thought to have been domesticated in both Mesoamerica and South America, but recent botanical and molecular studies indicate a South American origin²³, particularly a region of southwestern Brazil that supports a vegetation of seasonally moist tropical forest²⁰. Manioc pollen grains retrieved from two sites in Belize dated to 4,700 BP may represent the spread of this cultivar into Mesoamerica²⁴. Arrowroot's area of origin is unclear, but the seasonally moist forest of northern South America is a likely candidate²³. Our data indicate that prehistoric societies in tropical America domesticated and dispersed manioc and arrowroot by 7,000–6,000 BP, supporting an early transition from foraging to food production using root crops in the lowland Neotropical forest^{2,3,23,25}. □

Methods

Archaeological field research

Artefacts and sediments for starch analysis were retrieved while the excavations were in progress using starch-free plastic gloves, and immediately put into plastic bags. After artefacts were removed from the ground, sediments from directly beneath and around their periphery to a distance of 5–10 cm were sampled. Column samples, each measuring from 20 × 20 cm to 30 × 30 cm, were taken every 5 cm from profiled walls of squares after excavations were complete.

Laboratory research

Step 1: Unwashed grinding stones were examined under a stereoscopic microscope at a power of ×100. The point of a fine needle was inserted into cracks and crevices on the grinding facets, pounding edges and non-utilized surfaces of the artefacts to loosen and remove any residue. The residue was mounted in water on a slide and examined with polarized and unpolarized light at a power of ×400. 10–13 locations covering the utilized and non-utilized facets of tools were examined in this manner. Step 2: Tools were washed with a brush underneath running water, and starch was removed from the sediment using techniques described in detail elsewhere²⁶ (such as addition of a heavy metal liquid at a density of 1.8 made from a solution of caesium chloride, CsCl). Step 3: Washed tools were shaken in an ultrasound for five to ten minutes to further dislodge adhering sediment and starch, and starch was then isolated using CsCl, as in Step 2. Step 4: 10–20 cm³ of sediment sampled from beneath and around the peripheries of tools while the excavations were in progress was analysed for starch content, as in Step 2. Step 5: 24 sediment samples from two stratigraphic columns (from excavation squares 1N6E and 155E) were analysed for starch content as in Step 2. Four non-artefactual angular rocks that were a natural component of the site and closely associated sediment were sampled from the bottom to the top of the excavations and were analysed using steps 1–4. All archaeological starch grains possessed the starch-specific extinction cross when examined under cross-polarized light.

Starch grain identification

We used a modern reference collection of 170 species from 36 families of plants. It comprises: (1) most domesticated plants known or thought to have been used in pre-Columbian Panama and close wild relatives; (2) non-cultivated species used as foodstuffs (for example, Fabaceae, Marantaceae, various tree fruits); and (3) plants commonly used for medicinal, craft and ritualistic purposes in local indigenous economies today. Included were 12 different populations of manioc and two populations of *Dioscorea trifida* (domesticated yam) grown by local farmers, 15 races of Latin American maize, the putative wild ancestors of maize (*Zea mays* spp. *parviglumis*) and manioc (*Manihot esculenta* ssp. *flabellifolia*), and 48 wild species of Poaceae together with 12 wild species of *Dioscorea* from Panama. The numerous studies available on starch grain morphology were also consulted^{15–19,27}.

Our starch keys and classifications emphasize attributes demonstrated by these studies to be useful in identification: overall grain shape (spherical, bell-shaped, ovoid); contour

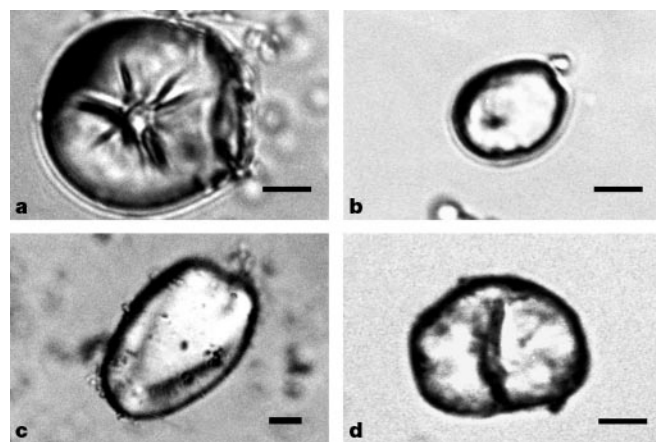


Figure 1 Various starch grains. **a**, A starch grain from manioc recovered from milling stone 42. **b**, A starch grain from the putative wild ancestor of manioc, *Manihot esculenta* ssp. *flabellifolia*. **c**, A starch grain from a species of *Dioscorea* recovered from milling stone 350. **d**, A starch grain from maize from milling stone 42. This type is unique to maize and is found in Race Jala from Mexico. Scale bar, 5 µm (**a**, **b**, **d**); 10 µm (**c**).

and surface features; position and form of the hila (the botanical centre of the grain) and fissure; number and characteristics of pressure facets; and presence or absence of demonstrable lamellae. Manioc, arrowroot, *Dioscorea* spp. and maize contribute unique, individual grains (Fig. 1a, c, d). To confirm identifications of manioc and maize we used attribute combinations in a multiple grain analysis, which is the most conservative means to distinguish species because it provides population signatures for specific species and takes into account intra- and inter-species variation in grain attributes^{15,16,19,27}.

Starch grains in roots of modern domesticated *Manihot esculenta* are compound and predominantly bell-shaped with smooth and highly transparent surfaces, one to five unbanded pressure facets with rounded edges, hila that are centric and open, and no demonstrable lamellae. Many grains have fissures that are often crossed, winged to y-shaped, or stellated (Fig. 1a). This combination of characteristics distinguishes manioc from all other taxa, including the few others with bell-shaped starches such as *Pachyrhizus* spp.¹⁸, in which grains have banded and angled pressure facets, closed or semi-open hila, and no or simple fissures. Our results indicate that bell-shaped grains are primarily confined to subterranean organs of plants. Starch grains from maize kernels are predominantly spherical and smooth (flour corns) or mostly angular and four- to five-sided with a rough, grooved surface (popcorns and flint corns). In this and other studies^{16,17,27}, endosperm type is thus shown to exert considerable influence on starch morphology in maize. Our morphological and size determinations for different species incorporated sampling of multiple reference points within a single specimen, to account for within-sample variability.

Received 13 June; accepted 14 August 2000.

- Harlan, J. R. *Crops and Man* 2nd edn (American Society of Agronomy, Crop Science Society of America, Madison, WI, 1992).
- Sauer, C. O. *Agricultural Origins and Dispersals* (American Geographical Society, New York, 1952).
- Lathrap, D. W. *The Upper Amazon* (Praeger, New York, 1970).
- Mangelsdorf, P. C. Review of *Agricultural Origins and Dispersals*. *American Antiquity* **19**, 87–90 (1953).
- Meggers, B. J. Environmental limitations on the development of culture. *Am. Anthropol.* **56**, 801–824 (1954).
- Smith, B. D. *The Emergence of Agriculture* (Scientific American, New York, 1998).
- Ranere, A. J. & Hansell, P. in *Prehistoric Coastal Adaptations* (eds Stark, B. L. & Voorhees, B.) 43–59 (Academic, New York, 1978).
- Wilding, L. P. Radiocarbon dating of biogenetic opal. *Science* **156**, 66–67 (1967).
- Kelly, E. F., Amundson, R. G., Marino, B. D. & Deniro, M. J. Stable isotope ratios of carbon in phytoliths as a quantitative method of monitoring vegetation and climate change. *Quat. Res.* **35**, 222–233 (1991).
- Ranere, A. J. & Cooke, R. G. in *Paths to Central American Prehistory* (ed. Lange, F. W.) 49–77 (Univ. Press of Colorado, Niwot, 1996).
- Ranere, A. J. in *Lithic Technology: Making and Using Stone Tools* (ed. Swanson, E.) 173–210 (Mouton, The Hague, 1975).
- Ranere, A. J. in *Adaptive Radiations in Prehistoric Panama* (eds Linares, O. F. & Ranere, A. J.) 118–137 (Peabody Museum Monographs No. 5, Harvard Univ., Cambridge, 1980).
- Cooke, R. G. in *The Emergence of Pottery: Technology and Innovation in Ancient Societies* (eds Barnett, W. K. & Hoopes, J. W.) 169–184 (Smithsonian Institution Press, Washington DC, 1995).
- Piperno, D. R. & Holst, I. The presence of starch grains on prehistoric stone tools from the lowland Neotropics: Indications of early tuber use and agriculture in Panama. *J. Archaeol. Sci.* **25**, 765–776 (1998).
- Reichert, E. T. *The Differentiation and Specificity of Starches in Relation to Genera, Species, Etc.* (Carnegie Institution of Washington, Washington, DC, 1913).
- Seidemann, J. *Stärke Atlas* (Paul Parey, Berlin, 1966).
- Snyder, E. M. in *Starch: Chemistry and Technology* (eds Whistler, R. L., Bemiller, J. N. & Paschall, E. F.) 661–673 (Academic, Orlando, 1984).
- Ugent, D., Pozorski, S. & Pozorski, T. Archaeological manioc (*Manihot*) from coastal Peru. *Econ. Bot.* **40**, 78–102 (1986).
- Ugent, D., Dillehay, T. & Ramirez, C. Potato remains from a Late Pleistocene settlement in south-central Chile. *Econ. Bot.* **41**, 17–27 (1982).
- Olsen, K. M. & Schall, B. A. Evidence on the origin of cassava: Phylogeography of *Manihot esculenta*. *PNAS* **96**, 5586–5598 (1999).
- Barton, H., Torrence, R. & Fullagar, R. Clues to stone tool function re-examined: comparing starch frequencies on used and unused obsidian artifacts. *J. Archaeol. Sci.* **25**, 1231–1238 (1998).
- Piperno, D. R., Bush, M. B. & Colinvaux, P. A. Paleocological perspectives on human adaptation in central Panama. II. The Holocene. *Geoarchaeology* **6**, 227–250 (1991).
- Piperno, D. R. & Pearsall, D. M. *The Origins of Agriculture in the Lowland Neotropics* (Academic, San Diego, 1998).
- Pohl, M. et al. Early agriculture in the Maya lowlands. *Latin Am. Antiquity* **7**, 355–372 (1996).
- Harris, D. R. The origins of agriculture in the tropics. *Am. Sci.* **60**, 180–193 (1972).
- Therin, R., Fullagar, R. & Torrence, R. in *The Prehistory of Food* (eds Gosden, C. & Hather, J.) 438–462 (Routledge, London, 1999).
- Cortella, A. R. & Pochettino, M. L. Starch grain analysis as a microscopic diagnostic feature in the identification of plant material. *Econ. Bot.* **48**, 171–181 (1994).

Acknowledgements

We thank L. Perry and M. Pohl for discussion and comments, C. Galdames for collecting economic plants, and K. Olsen for providing wild manioc. This work was supported by the Smithsonian Tropical Research Institute (STRI) and a grant to the STRI from the Andrew W. Mellon Foundation.

Correspondence and requests for materials should be addressed to D.R.P. (e-mail: pipernod@stri.org).

Isolation of a 250 million-year-old halotolerant bacterium from a primary salt crystal

Russell H. Vreeland*, William D. Rosenzweig* & Dennis W. Powers†

* Department of Biology, West Chester University, West Chester, Pennsylvania 19383, USA

† Consulting Geologist, Box 87, Anthony, Texas 79821, USA

Bacteria have been found associated with a variety of ancient samples¹, however few studies are generally accepted due to questions about sample quality and contamination. When Cano and Borucki² isolated a strain of *Bacillus sphaericus* from an extinct bee trapped in 25–30 million-year-old amber, careful sample selection and stringent sterilization techniques were the keys to acceptance. Here we report the isolation and growth of a previously unrecognized spore-forming bacterium (*Bacillus* species, designated 2-9-3) from a brine inclusion within a 250 million-year-old salt crystal from the Permian Salado Formation. Complete gene sequences of the 16S ribosomal DNA show that the organism is part of the lineage of *Bacillus marismortui* and *Virgibacillus pantothenicus*. Delicate crystal structures and sedi-

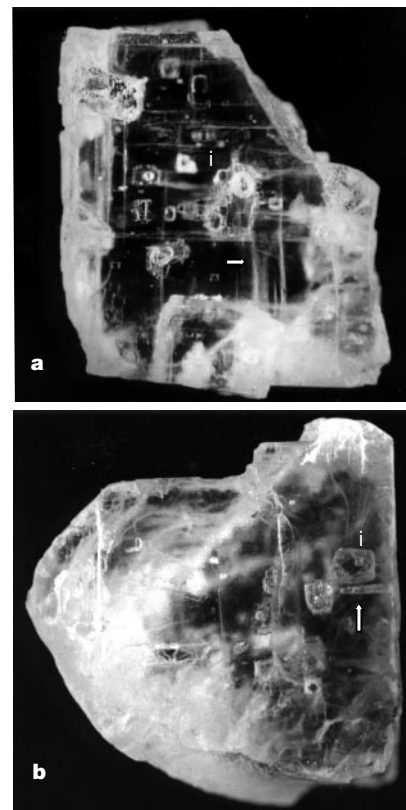


Figure 1 Halite crystals taken from the dissolution pipe at the 569 m (1850 ft) level of the Salado Formation (air intake shaft¹⁸ for the Waste Isolation Pilot Plant, Carlsbad, New Mexico) in October 1998. **a**, This crystal was rejected for sampling. The crystal has numerous fracture lines (arrow) and misshapen inclusions with gas bubbles (i). This crystal measured 3.2 × 2.8 × 0.9 cm. **b**, The sample that yielded *Bacillus* strain 2-9-3. This crystal measured 3.5 × 3.5 × 2.5 cm. The inclusion that contained the bacterium (below i) measured approximately 3 × 3 × 1 mm (9 cubic mm). The drill hole made during sampling of inclusion B (Table 1) is visible above the arrow. The thickness of this crystal obscures some of the internal details (the drill hole for inclusion A) in the photograph.

mentary features indicate the salt has not recrystallized since formation. Samples were rejected if brine inclusions showed physical signs of possible contamination. Surfaces of salt crystal samples were sterilized with strong alkali and acid before extracting brines from inclusions. Sterilization procedures reduce the probability of contamination to less than 1 in 10^9 .

Bacteria associated with ancient samples include *Halococcus salifodinae*³, and several other unnamed halophilic Archaea isolated from salt formations^{4,5}. There are a variety of claims of isolation of eubacteria from salts (reviewed in ref. 1). Nevertheless, the most acceptable studies are of bacteria isolated from amber^{2,6,7}. Before the ancient age of any organism can be firmly established and accepted, two basic criteria should be met: the geological age of the formation or period of isolation should be established; and clear laboratory data should show steps taken to maintain a low probability of contamination^{1,8,9}.

The Salado is generally accepted to be at least 250 Myr old. This age is supported by invertebrate fossils^{10,11} and radiometric ages¹² from overlying formations, as well as radiometric ages of langbeinite minerals¹³ taken directly from the Salado. Crystal features and original structures in the Salado beds are evidence of integrity over the last 250 Myr. Delicate fluid inclusions commonly trapped by halite crystals as they grow are one indicator that halite has not recrystallized since being deposited^{8,14–20}. The Salado also includes dissolution pits and pipes that are filled with coarse, clear halite with fewer, but larger, fluid inclusions^{18,20,21}. These pits and pipes are overlaid by undisturbed halite beds, signifying that they are the same age (Late Permian) as the formation^{18,20,21}. These crystals also do not show boundary relationships of recrystallized halite¹⁵ that would indicate a younger age. Because fluid inclusions in these crystals are much larger than in other primary halite, they are preferred for the type of sampling performed here (Fig. 1).

It is fair to note that the radiometric age for many potassium-bearing minerals in the Salado range from 210–215 Myr²². However, the ⁴⁰Ar–³⁹Ar plateau age for langbeinite crystals taken directly from the Salado is reproducibly 251 Myr²². The cause of the difference between the evident age of deposition and radiometric age is undetermined²². Although it could be argued that these younger dates represent a last stage of mineralization (and closing of the system at 210 Myr ago), textural data^{18,20,21}, langbeinite dates¹³ and crystals of KCl in fluid inclusions in primary halite crystals¹⁹ are strong evidence that our samples remained isolated for 250 Myr, not 210 Myr.

The crystal samples were taken using a hand-held electric drill and coring bit. The underground mining techniques used in this facility do not include blasting and the shaft was constructed by upreaming¹⁸. This limits mechanical damage to the wall rock compared with blasting. Only crystals from primary sedimentary features were removed. Each crystal used for microbiological sampling was examined, and crystals were rejected as damaged if inclusions showed rounded edges, elongation or stress-related

roughness along the walls of the inclusion¹⁵. A large number of crystals were rejected, but this procedure helps ensure that inclusions were not contaminated during collection. The 250-Myr-old crystal sample that contained viable bacteria was found in a dissolution pipe 569 m below the surface, in the wall of the air intake shaft of the Waste Isolation Pilot Plant, Carlsbad, New Mexico (Fig. 1). The crystal measured $3.5 \times 3.5 \times 2.5$ cm and was one of two primary crystals from this site that contained living bacteria (Table 1).

In addition to establishing that the crystals are primary, sterility is critical when attempting to find organisms trapped in the precipitating crystals^{1,8,9}. The sample surface must be sterilized by a verifiable technique, and the sample must be extracted under aseptic conditions. The surface of samples like amber can be sterilized using glutaraldehyde, bleach and ethanol^{2,6,7}. Unfortunately, the reactivity and solubility of NaCl means this type of sterilization cannot be applied to these rocks.

Recently, several new bacterial species and/or genera were cultured from unsterilized Salado halite dissolved in commonly used halophile medium²³. The largest bacterial population found in the Salado proved to be less than 1×10^5 colony-forming units (c.f.u.) per gram of salt.

Immersing crystals in 10M NaOH, washing, then immersing in 10M HCl provides a minimal sterility assurance level of at least 1 in 10^{-11} (ref. 24). Steam sterilization of all extraction equipment (see Methods) provides a sterility assurance level of 1 in 10^{-9} . Consequently, we assume that 1 in 10^{-9} is the overall minimum sterility assurance level for the experiment. In addition to the sterility assurances, contamination controls were used during all drilling operations. These included open plates of media distributed at random in the hood, streaking every sterilized crystal face before drilling, and an open plate of medium placed to catch the salt fines produced by the drill bits. These controls have been consistently negative for all salt crystals sampled.

The inclusion (crystal 1850-3A, Table 1) that yielded strain 2-9-3 measured about $3 \times 3 \times 1$ mm ($\sim 9 \mu\text{l}$ of fluid). This fluid was inoculated into two different media: caesin-derived amino acids medium (CAS)²⁵ and glycerol–acetate medium (GA) (Table 1) (R.H.V., W.D.R. & M. Kamekura, manuscript in preparation). Only the CAS enrichment yielded any bacteria. Streak plates made directly from the enrichment tube yielded only one type of colony. Another bacterial strain was present in a second inclusion (1850-3C) in this same crystal. A third strain was present in crystal 1850-2. Both of these strains are currently being compared with isolate 2-9-3 and other bacilli. The *Bacillus* species (designated 2-9-3) was first isolated on CAS agar supplemented with 20% (w/v) NaCl. The original growth conditions were 37°C with aeration on a rotating drum.

The complete 16S rDNA sequence of 2-9-3 (GenBank accession number AF166093) was compared with all other known rDNA sequences through a GenBank BLAST search. Similar rDNA sequences were downloaded from the database Internet site using GenBank accession numbers and aligned using ClustalW 1.7. The phylogeny of the new isolate was then determined using programs in

Table 1 Crystal sampling within the 569 m (1850 ft) region of the Salado Formation

Crystal number	Inclusion size (mm)	Fluid recovered (μl)	Medium type*		
			CAS	CG	GA
1850-1	$2 \times 3 \times 3$	2	–	–	n.i.†
1850-2	$2.5 \times 3 \times 4.5$	10	+	–	–
1850-3A‡	$3 \times 3 \times 1$	9	+	n.i.	–
1850-3B	$3 \times 2 \times 2$	10	–	–	n.i.
1850-3C	$4 \times 3 \times 3.5$	12	+	–	n.i.

All of the crystals listed were sampled on the same day. Sixty-one additional inclusions in 50 crystals have also been sampled on additional days. A total of 365.5 μl of brine fluid have been inoculated into over 70 different test tubes. To date, only the three tubes listed have produced growth.

* CAS, caesin-derived amino acids medium; CG, cellulose glucose medium²⁵; GA, glycerol–acetate medium.

† n.i., none inoculated.

‡ A, B and C refer to different inclusions within crystal-3.

Table 2 Contamination control: sterilization procedures eliminate the 250-Myr-old isolate from crystal-sampling equipment

2-9-3-contaminated materials	Pre-sterilization		Post-sterilization	
	Day 3	Day 7	Day 3	Day 7
Stock culture	+	n.d.†	–‡	–
21 drill bits	+	n.d.	–	–
4 plastic forceps	+	n.d.	–	–
4 metal forceps	+	n.d.	–	–
5-microlitre syringes	+	n.d.	–	–

* 100% of plates positive for growth.

† n.d., not determined as all tests were positive at day 3.

‡ 100% of plates negative for growth.

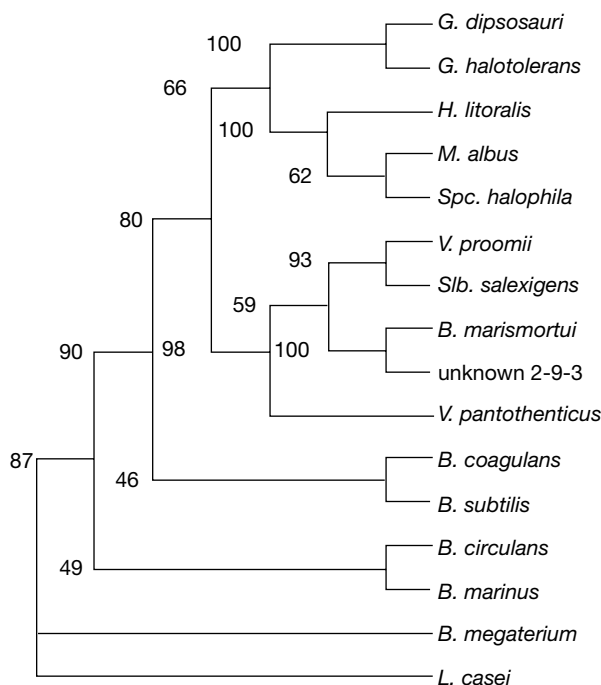


Figure 2 Phylogenetic tree developed from the 16S rDNA sequence of unknown 2-9-3 compared with several other species of the family *Bacillaceae*. The bootstrap values from 100 repetitions for each grouping are shown to the left of the branch. The Genbank accession numbers for the organisms used in this analysis are *Bacillus megaterium* X60629; *B. subtilis* X60646; *B. marismortui* AJ009793; isolate 2-9-3 AF166093; *B. coagulans* X60614; *B. circulans* X60613; *B. marinus* AJ237708; *Virgibacillus pantothenticus* X60627; *V. proomii* AJ0122667; *Halobacillus littoralis* X94558; *Gracilbacillus halotolerans* AF036922; *G. dipsosauri* X82436; *Salibacillus salexigens* (Slb.) Y11603; *Sporosarcina halophila* (Spc.) X62174. *Lactobacillus casei* D16553 served as the outgroup.

the Phylip phylogenetic software package (<http://evolution.genetics.washington.edu/phylip.html>). These analyses showed that the organism was most similar to *Bacillus marismortui* (99% similarity (S)²⁶ and *Virgibacillus pantothenticus* (97.5% S)²⁷. Phylogenetic analysis showed that isolate 2-9-3, *B. marismortui*²⁶, *V. pantothenticus*²⁷, *Salibacillus salexigens*^{28,29} and *V. proomii*²⁷ form a distinct lineage within the larger *Bacillus* cluster (Fig. 2).

The sterilization technique that we used²⁴ was initially developed using a non-spore-forming, Gram-negative, halotolerant eubacterium and a halophilic Archeon. As the first isolate found with these techniques proved to be a spore-forming, Gram-positive *Bacillus*, re-testing the procedures using 2-9-3 seemed warranted. Exposing a culture of 2-9-3 containing 10^8 cells and spores to 10 M NaOH for 1 min reduced the population by at least 10^3 c.f.u. ml⁻¹. Exposing the culture to 10 M HCl for 1 min reduced the number of viable cells and spores to undetectable levels (a reduction greater than 10^8 c.f.u. ml⁻¹). In addition, all sampling equipment was first contaminated with cells and spores of 2-9-3 then autoclaved. The testing showed that the sterilization cycle used for this work eliminated the organism (Table 2). None of the plates exposed in the laminar flow hood for 50 min, or those used to sample the disinfected clamp used to hold crystals for drilling, produced viable colonies following 7 days of incubation.

As the crystal containing 2-9-3 was exposed to the 10 M NaOH for 5 min, washed in sterile brine, then exposed to 10 M HCl for 5 min, we believe that the crystal experienced the full effect of both sterilizing agents. Consequently, the probability that 2-9-3 was a surface contaminant on the crystal is less than 1 in 10^{-11} . The data in Table 2 and that of Rosenzweig *et al.*²⁴ show that this bacterium could not have survived as a contaminant on any equipment or

crystal used during this isolation. We conclude that strain 2-9-3 was isolated from the fluid inclusion contained within the 250-Myr-old primary crystal. The crystal sampled showed no sign of having been damaged or penetrated before this work. There is strong support for an age of 250 Myr for the crystal. Therefore, we believe that strain 2-9-3 is a bacterium that was present in a hypersaline environment during the late Permian, trapped inside a crystal at that time and survived within the crystal until the present.

The distribution of living microbes in salt crystals is not known at present. There may be regions of these ancient formations containing significant numbers of live bacteria²³. Specific crystals may harbour several inclusions containing live organisms, although other inclusions and/or crystals from the same sample appear sterile (Table 1, ref. 23). However, only three enrichments grew from inclusions in 2 crystals out of 53 crystals sampled so far, indicating that survival may be a rare occurrence. At this point, we cannot address the mechanisms of survival, which may require a complex interaction between abiotic conditions in specific inclusions and biotic capability. In the case of *Bacillus* 2-9-3, sporulation would be a potential mechanism of survival as high ionic strength media does stimulate spore formation *in vitro*. However, we have no direct evidence of the presence of spores in the brine extracted from the inclusions.

The presence of 2-9-3 in these crystals is consistent with the basic sedimentary patterns in the Salado. Dissolution pits show that less saline waters covered exposed salt beds, dissolved salt to form pipes and began another round of deposition. The water with a lower salt content probably carried a variety of microorganisms into the evaporating salt pan. Bacteria with an ability to adapt to the increasing salt survived, and a bacterial spore in the brine may have been trapped in the crystals filling the pipes. Once the organism was fully encased, evolutionary pressures would have been relieved. Strains remaining in the standing brine may have continued along divergent evolutionary pathways.

Whether or not *Bacillus* strain 2-9-3 is an ancestor to these other organisms cannot be determined through the study of a single highly conserved gene such as the 16S rDNA³⁰. It is interesting to note that there are 45 unequivocal sequence differences (in 1,560 base pairs) existing between the 16S rDNA of 2-9-3 and *Virgibacillus pantothenticus*. Assuming that these two organisms have been separated since 2-9-3 was trapped in the crystal (250 Myr), this becomes a base-pair substitution rate equal to 1×10^{-10} substitutions per site per year. This substitution rate is slower than that found for *Bacillus sphaericus*² but identical to that calculated for enteric microbes² and Aphid symbionts². The reasons for minimal sequence differences between 2-9-3 and *Bacillus marismortui* are unknown.

Previous reports of bacteria isolated from salt have met with resistance, partly because of concerns for contamination. Samples from many studies have been described as brines, rock salt, efflorescences, recrystallized salt or wet salt. These terms provide little or no assurance of the provenance of the sample or geological history. Flowing brines and salt efflorescences, for example, may be contaminated and young, and recrystallized salts may be of indeterminate age. In our study, however, we support the discovery by geological evidence of the overall age and stability of both the formation and the individual crystals, as well as the documented sterility maintained during the sampling and lab work. This research doubles the length of time bacteria can survive trapped inside stable geological materials. □

Methods

Once selected, crystals were gently cleaned with a rotary grinder, and polished using distilled water. This process removed any adhering soils and eliminated small fissures that might prevent the sterilant from reaching all portions of the crystal surface. The polishing step also increased the ability to see the inclusion and monitor the progress of the drill bit. The cleaned crystals were moved into a Biological Level 3 clean-room facility physically separated from our main laboratory. Personnel working with crystals wore clean room

coveralls, boots, hats and latex gloves. Drill bits, media, washes and syringes were sterilized by autoclaving. All autoclaving was carried out for 40 min at 121 °C in a double door pass through autoclave, continually serviced under annual maintenance contracts with quarterly inspection and testing. Drill bits, forceps and syringes used for samplings were individually packaged before sterilization. All media were placed into test tubes (no more than 8 ml per tube) and autoclaved. The manufacturer's specifications indicate that all of the autoclaved materials were sterilized to a sterility assurance level equal to 1×10^{-9} , or 1 chance of contamination in 10^9 . In addition, all media to be used for samples were placed into an incubator at 37 °C for 3 days before use. If any medium batch contained even a single contaminated tube, the entire batch was discarded.

Crystals were sterilized as described²⁴. All sterilants and brine washes were placed into sterile covered beakers. The cleaned crystals were immersed in 10 M NaOH for 5 min, washed in sterile saturated salt brine for 2 min, then immersed in 10 M HCl for 5 min. Following the HCl, each crystal was washed in sterile saturated salt brine buffered with sodium carbonate. This protected the stainless steel of the laminar flow hoods and drill bits from the HCl. All washes and sterilants were changed after each crystal. Sterilized crystals were removed from the buffered brine inside a dedicated Class IIA laminar flow hood fitted with an HEPA filter. The laminar flow hood was disinfected with germicidal ultraviolet light for 2 h before use. All hood surfaces and non-autoclavable equipment were disinfected with a commercial disinfectant between each drilling²⁴. Sterile crystals were tightly held in a pinch clamp and penetrated with a sterile 0.5 mm wire drill operated with a micromanipulator. The inclusion samples were extracted from the crystals using sterile 25- or 250- μ l syringes. The brine samples were inoculated into the sterilized growth medium.

To be sure that the spore-forming isolate was not a contaminant, the various materials used in the handling, drilling and extraction of fluids from salt crystals were deliberately contaminated with a 36-h culture of isolate 2-9-3 containing a mixture of cells and spores. All contaminated materials were streaked onto plates of tryptic-soy agar (TSA), and CAS agar amended with either 8% (CAS-8) or 20% (CAS-20) NaCl. One plate of each medium was used. All materials were then packaged as normal for crystal work, autoclaved and, after cooling, were again streaked onto TSA, CAS-8 and CAS-20 plates. All plates were incubated at 37 °C and scored for the presence of growth after 3 and 7 days. The clamp used to hold crystals during drilling in the biosafety hood, was disinfected with Wescodyne and streaked onto plates of TSA, CAS-8 and CAS-20 medium. Two plates each of TSA, CAS-8 and CAS-20 were opened and placed into the running biosafety hood following 1 h of ultraviolet light. The plates were exposed for at least 50 min, equal to the time needed to sample three crystals. The stock culture of 2-9-3 used for contamination was also streaked onto TSA, CAS-8 and CAS-20 medium before and after autoclaving.

Received 15 November 1999; accepted 4 July 2000.

- Grant, W. D., Gemmell, R. T. & McGenity, T. J. Halobacteria: the evidence for longevity. *Extremophiles* **2**, 279–287 (1998).
- Cano, R. J. & Borucki, M. Revival and identification of bacterial spores in 25 to 40 million year old Dominican amber. *Science* **268**, 1060–1064 (1995).
- Denner, E. B. M. *et al.* *Halococcus salifodinae* sp. nov., an archaeal isolate from an Austrian salt mine. *Int. J. Syst. Bacteriol.* **44**, 774–780 (1994).
- Huval, J. H. & Vreeland, R. H. in *General and Applied Aspects of Halophilic Bacteria*. Vol. 201 (ed. Rodriguez-Valera, F.) 53–62 (Plenum, New York, 1991).
- Norton, C. F., McGenity, T. J. & Grant, W. D. Archaeal halophiles (halobacteria) from two British salt mines. *J. Gen. Microbiol.* **139**, 1077–1081 (1993).
- Greenblatt, C. L. *et al.* Diversity of microorganisms isolated from amber. *Microb. Ecol.* **38**, 58–68 (1999).
- Lambert, L. H. *et al.* *Staphylococcus succinus* sp. nov., isolated from Dominican amber. *Int. J. Syst. Bacteriol.* **48**, 511–518 (1998).
- Vreeland, R. H. & Powers, D. W. in *Microbiology and Biogeochemistry of Hypersaline Environments* (ed. Oren, A.) 53–74 (CRC, Boca Raton, Florida, 1999).
- Vreeland, R. H. & Rosenzweig, W. D. in *Enigmatic and Extreme Microorganisms*. (ed. Seckbach, J.) 387–398 (Kluwer, Delft, 1998).
- Croft, J. S. Upper Permian conodonts and other microfossils from the Pinery and Lamar Limestone Members of the Bell Canyon Formation and from the Rustler Formation, west Texas. Thesis, Ohio State Univ. (1978).
- Walter, J. C. Paleontology of the Rustler Formation, Culberson County, Texas. *J. of Paleontol.* **27**, 679–702 (1953).
- Renne, P. R., Steiner, M. B., Sharp, W. D., Ludwig, K. R. & Fanning, C. M. 40/39 Ar and U/Pb SHRIMP dating of latest Permian tephra in the Midland Basin Texas. *EOS* **77**, 794 (1996).
- Renne, P. R., Sharp, W. D. & Becker, T. A. ⁴⁰Ar/³⁹Ar dating of langbeinite [K₂Mg₂(SO₄)₃] in late Permian evaporites of the Salado Formation, Southeastern New Mexico, USA. *Mineral. Mag.* **62A**, 1253–1254 (1998).
- Hardie, L. A., Lowenstein, T. K. & Spencer, R. J. The problem of distinguishing between primary and secondary features in evaporites. *Sixth Int. Symp. On Salt* **1**, 11–39 (1983).
- Roedder, E. The fluids in salt. *Amer. Mineral.* **69**, 413–439 (1984).
- Lowenstein, T. K. & Hardie, L. A. Criteria for recognition of salt-pan evaporites. *Sedimentol.* **32**, 627–644 (1985).
- Lowenstein, T. K. Origin of depositional cycles in a Permian "saline giant": the Salado (McNutt zone) evaporites of New Mexico and Texas. *Geol. Soc. Am. Bull.* **100**, 592–608 (1988).
- Holt, R. M. & Powers, D. W. Geological mapping of the air intake shaft at the Waste Isolation Pilot Plant. Report no. DOE/WIPP 90-051, 1–90 (U. S. Department of Energy, Carlsbad NM, 1990).
- Lowenstein, T. K. & Spencer, R. J. Syndepositional origin of potash evaporites: petrographic and fluid inclusion evidence. *Am. J. Science* **290**, 1–42 (1990).
- Holt, R. M. & Powers, D. W. in *Geological and Hydrological Studies of Evaporites in the Northern Delaware Basin for the Waste Isolation Pilot Plant (WIPP)*, New Mexico (eds Powers, D. W., Holt, R. M., Beauheim, R. L. & Rempe, N.). *Geol. Soc. Am. Guidebook* **14**, 45–78 (1990).
- Powers, D. W. & Hassinger, B. W. Synsedimentary dissolution pits of halite of the Permian Salado Formation, southeastern New Mexico. *J. Sed. Petrol.* **55**, 769–773 (1985).

- Onstott, T. C., Mueller, C., Mikulski, K., Vicenzi, E. & Powers, D. W. ⁴⁰Ar/³⁹Ar laser microprobe dating of polyhalite from bedded, late Permian evaporites. *EOS* **76**, S285 (1995).
- Vreeland, R. H., Piselli, A. F., McDonough, S. & Myer, S. S. Distribution and diversity of halophilic bacteria in a subsurface salt formation. *Extremophiles* **2**, 321–331 (1998).
- Rosenzweig, W. D., Peterson, J., Woish, J. & Vreeland, R. H. Development of a protocol to retrieve microorganisms from ancient salt crystals. *Geomicrobiol.* (in the press).
- Vreeland, R. H., Anderson, R. & Murray, R. G. E. Cell wall and phospholipid composition and their contribution to the salt tolerance of *Halomonas elongata*. *J. Bacteriol.* **160**, 879–883 (1984).
- Arahal, D. R., Marquez, M. C., Volcani, B. E., Schleifer, K. H. & Ventosa, A. *Bacillus marismortui* sp. nov., a new moderately halophilic species from the Dead Sea. *Int. J. Syst. Bacteriol.* **49**, 521–530 (1999).
- Heyndrickx, M. *et al.* *Virgibacillus*: a new genus to accommodate *Bacillus pantothenicus* (Proom and Knight 1950). Emended description of *Virgibacillus pantothenicus*. *Int. J. Syst. Bacteriol.* **48**, 99–106 (1998).
- Garabito, M. J., Arahal, D. R., Mellado, E., Marquez, M. C. & Ventosa, A. *Bacillus salaxigens* sp. nov., a new moderately halophilic *Bacillus* species. *Int. J. Syst. Bacteriol.* **47**, 735–741 (1997).
- Waino, M., Tindall, B. J., Schumann, P. & Ingvorsen, K. *Gracilibacillus* gen. nov., with description of *Gracilibacillus halotolerans* gen. nov. sp. nov.; transfer of *Bacillus dipsosauri* to *Gracilibacillus dipsosauri* comb. nov., and *Bacillus salaxigens* to the genus *Salibacillus* gen. nov., as *Salibacillus salaxigens* comb. nov. *Int. J. Syst. Bacteriol.* **49**, 821–831 (1999).
- Courmoyer, B. & Lavire, C. Analysis of *Frankia* evolutionary radiation using *glnII* sequences. *FEMS Microbiol. Lett.* **177**, 29–34 (1999).

Acknowledgements

The authors acknowledge the following people who helped obtain the crystal samples for this research: D. Belski, N. Rempe, R. Carrasco, T. Garcia, D. Acevedo, S. Britain, E. Keyser, B. Kinsall, A. Morin and T. Padilla. This research was supported by the US National Science Foundation: Life in Extreme Environments Program (EAR Lexen).

Correspondence and requests for materials should be addressed to R.H.V. (e-mail: rvreeland@wcupa.edu).

The proteins of linked genes evolve at similar rates

Elizabeth J. B. Williams & Laurence D. Hurst

Department of Biology and Biochemistry, University of Bath, Claverton Down, Bath, BA2 7AY, UK

Much more variation in the rate of protein evolution occurs than is expected by chance¹. But why some proteins evolve rapidly but others slowly is poorly resolved. It was proposed, for example, that essential genes might evolve slower than dispensable ones², but this is not the case³; and despite earlier claims⁴, rates of evolution do not correlate with amino-acid composition⁵. A few patterns have been found: proteins involved in antagonistic co-evolution (for example, immune genes^{3,6}, parasite antigens⁷ and reproductive conflict genes^{8–10}) tend to be rapidly evolving, and there is a correlation between the rate of protein evolution and the mutation rate of the gene^{1,11,12}. Here we report a new highly statistically significant predictor of a protein's rate of evolution, and show that linked genes have similar rates of protein evolution. There is also a weaker similarity of rates of silent site evolution (see ref. 13), which appears to be, in part, a consequence of the similarity in rates of protein evolution. The similarity in rates of protein evolution is not a consequence of underlying mutational patterns. A pronounced negative correlation between the rate of protein evolution and a covariant of the recombination rate indicates that rates of protein evolution possibly reflect, in part, the local strength of stabilizing selection.

To examine the effects of linkage on rates of evolution, we established a data set of rates of evolution at both non-synonymous (K_a) and synonymous (K_s) sites of mouse–rat orthologues with confirmed genomic location. Rates of evolution of genes were compared with those less than 1 centiMorgan (cM) apart, this being equivalent to on average about 2,000 kilobases (kb) in the mouse genome¹⁴. We calculated the modular difference (ΔK) between the K values (that is, either K_a or K_s , or K_a/K_s depending

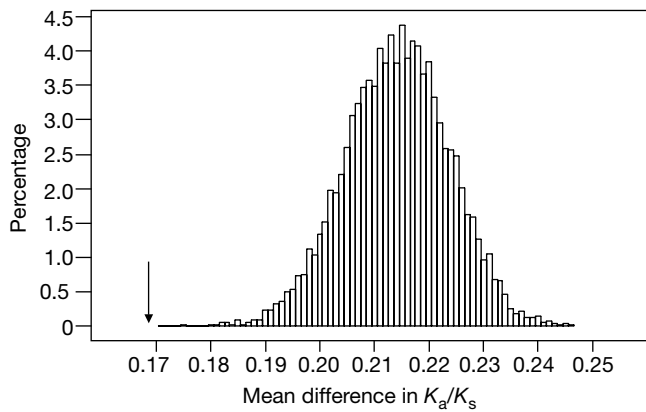


Figure 1 The distribution of 10,000 mean modular differences for K_a/K_s data. Each of the 10,000 values is the mean for a randomized version of the real K_a/K_s values, allowing in each 176 pairs of values. The actual mean modular difference for loci less than 1 cM apart is shown by an arrow.

on the analysis) for each of 176 pairs of linked genes (from 266 genes), and then determined the average.

To analyse the significance of this mean value, we followed a randomization method. We produced for each analysis 10,000 randomized data sets. In each random data set, we took all the K values from the actual data set and re-allocated them as pairs at random. For each randomized set, we calculated a mean modular difference in rate between the pairs. The proportion of the 10,000 random data sets with lower mean modular difference is a direct estimate of the P value that can be attached to the hypothesis that linked genes have similar rates of evolution.

With respect to the absolute rate of protein evolution (K_a) in 10,000 randomized data sets, none has a lower mean modular difference than that found in the real data set ($P < 0.0001$). There is also a local similarity of synonymous rates of evolution: only 163 of 10,000 random data sets have lower mean modular difference ($P = 0.016$). If the synonymous rate reflects the mutation rate, as often believed, this suggests that there may be differences between genomic regions in the underlying mutation rate (see also ref. 15). The data set is purged of tandem duplicates so neither concerted evolution nor genic non-independence can explain these patterns.

As a gene's mutation rate is likely to determine, in part, its rate of protein evolution (which explains, in part, the K_a by K_s correlation¹²), it is possible that the correlation in the rate of protein evolution of linked genes is a result of underlying mutational patterns alone. Comparing modular differences in the rate of protein evolution controlling for underlying mutation rate differences (that is, K_a/K_s), however, we find that of 10,000 random data sets, none shows a mean modular difference lower than (or even close to) that of the real data ($P < 0.0001$) (Fig. 1). The similarity of K_a/K_s of linked genes can also be demonstrated by comparing the distribution of the real 176 $\Delta K_a/K_s$ values with that from the randomized data (Fig. 2). This demonstrates an excess of small differences and a dearth of large differences in the real data (Mann–Whitney U -test: $P < 0.0001$).

By contrast, the synonymous rate bands can be explained in part as a mutational consequence of the protein rate bands. A non-synonymous mutation may be associated with a silent mutation at the adjacent site, for mechanistic reasons¹⁶. The spread of the non-synonymous mutation (by drift or selection) then takes the silent mutation at the adjacent site with it. Such tandem substitutions explain much of the covariation of K_a and K_s (ref. 12). To examine whether tandem substitutions might explain the synonymous rate domains, we removed them from all aligned sequences and re-calculated K_a and K_s . Although their removal reduces the mean number of synonymous sites per gene by only 3%, we find that the

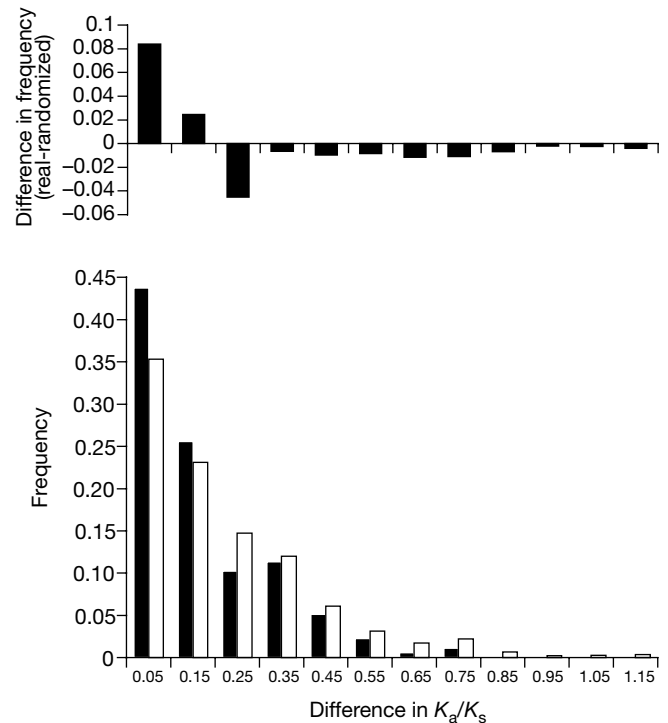


Figure 2 The frequency of differences in K_a/K_s between linked genes. Lower panel indicates the frequency of differences in K_a/K_s in 176 pairwise comparisons between linked genes (black bars), and in randomized data (white bars). Upper panel indicates the difference between the frequencies in the real and the randomized data. Numbers on the x axis refer to the median value of that bin (that is, 0.15 refers to values between 0.1 and 0.2).

synonymous rate domains disappear ($P = 0.08$), while the protein rate domains remain very evident ($P < 0.0001$).

Tandem substitutions need not be the only factor contributing to the pattern. As methylation density varies around the genome¹⁷, and as methylated CpG dinucleotides mutate about 10–20 times faster than other sites¹⁸, we also thought that the K_s similarity may be a result of differences in methyl-induced mutation rates. This may be the case, but the effect is not profound: after removal of CpG→TpG mutations at silent sites, the local correlation of silent rates of genes less than 1 cM apart is weaker but not greatly so ($P = 0.05$ rather than $P = 0.016$). Only 2% of silent sites were removed in this analysis.

The murid genome therefore appears to be split into protein rate domains, in which the proteins of linked genes have rates of evolution that are much more similar than expected by chance, even controlling for mutation rate differences. We estimate these rate domains to be only very few cM long; we base this assertion on the fact that when we compare genes that are between 1 and 2 cM apart the genes are no longer significantly similar in terms of K_a/K_s ($P = 0.10$). This value is based on 92 pairwise comparisons between 153 genes. Were the rate bands of the order of ~ 10 cM, then comparing genes 1–2 cM apart would still mostly be comparing genes within a band and so should have revealed P values of the order of those obtained in the less than 1-cM class.

What might explain the existence of protein rate domains? We propose two hypotheses. First, genes of comparable function tend to have comparable rates of evolution; for example, immune genes tend to be fast evolving, neurone-specific genes tend to be slow evolving³. If functionally similar genes tend to cluster, then this might explain the pattern. We shall leave a test of this hypothesis to future work.

Second, we thought that these domains might reflect differences

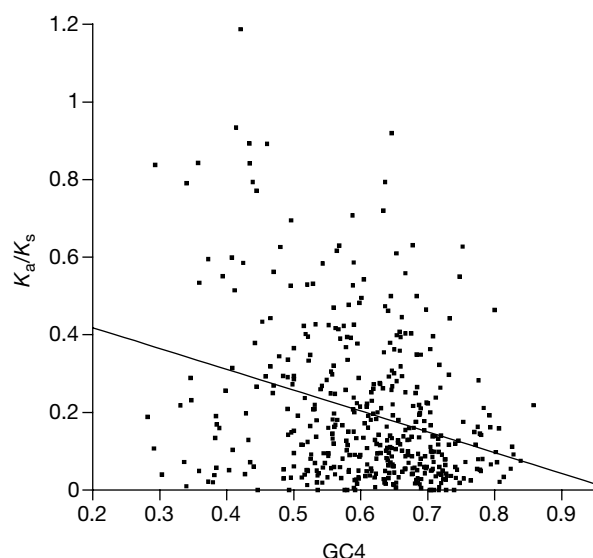


Figure 3 The covariance of the rate of protein evolution, controlling for the underlying mutation rate, and the GC4 content in the exons of the genes: $K_d/K_s = 0.53-0.54$ GC4, $R^2 = 0.09$, $P < 0.0001$.

in the strength of stabilizing selection. The local strength of stabilizing selection covaries with the local recombination rate¹⁹, which varies considerably around the mouse genome²⁰. Conveniently, the G + C content at fourfold redundant sites (GC4) is also thought to correlate with the recombination rate in mice²¹. This we can confirm. If recombination and G + C content are related, then we expect the Y chromosome to have a low G + C content (it never recombines), the X chromosome to have a higher one (it recombines in females only) and the autosomes to have the highest percentage of G + C. The Y chromosome has a mean G + C content at degenerate sites of 34% ($n = 7$), the X of 51% ($n = 37$) and the autosomes 61% ($n = 433$). These values are all highly significantly different from each other (Mann-Whitney U -test: $Y < X$, $P = 0.0005$; $X < A$, $P < 0.00001$). The values for the X and Y chromosomes cannot be explained as a correlate to hemizygous expression as imprinted genes ($n = 15$) have a G + C composition comparable to that of autosomal genes at 64%.

If, then, the local similarity of rates is due to variation in the intensity of stabilizing selection around the genome, both K_a and K_d/K_s should negatively correlate with GC4. We find a pronounced negative correlation between GC4 and both K_a and K_d/K_s for autosomal genes ($P < 0.0001$, $R^2 = 0.09$, $n = 431$) (Fig. 3). However, there may also be a difference in the sorts of genes in G + C-rich and G + C-poor isochores²², which might in part explain the pattern.

Our results contrast with those from a previous study that examined the local similarity of rates of evolution of genes. Comparing human with rodent sequences¹³, it was reported that the K_s of linked genes was highly similar, but that the similarity of K_a was of marginal significance. When we apply our methods to 176 randomly selected pairs in this rodent-human comparison, we find the same qualitative pattern, that is, strong local similarity in K_s ($P = 0.0013$) but not in K_a ($P = 0.02$). The discrepancy therefore appears to be neither due to differences in randomization methodology nor sample size. The greater distance in the rodent-human comparison as compared with the more recent mouse-rat split might possibly underlie this discrepancy.

The protein rate bands are not the only sort of mosaicism seen in the murid genome. The genome is, for example, split into G/R bands—regions that appear to correspond to late and early replicating DNA²³. Furthermore, the mammalian genome is a mosaic of isochores, blocks of DNA within which the proportion of the bases

G and C at non-coding sites (introns, third positions in codons, intergene spacer) is fairly uniform^{22,24}. However, murids do not have such a well-defined isochore structure²⁵. Nonetheless, considering the genes 1 cM apart, only 16 in 10,000 random data sets have lower mean modular difference in GC4 than the real linked genes ($P = 0.0016$), indicating that isochores do still exist (see also ref. 13).

Isochores and protein rate bands may be different things. Whereas the protein rate bands are not evident in genes between 1 and 2 cM ($P = 0.1$ compared with $P < 0.0001$ at less than 1 cM), the GC4 similarity of the same genes at this distance remains almost as profound as between 0 and 1 cM ($P = 0.0037$ rather than $P = 0.0016$). This also weakly indicates that isochores may be longer than usually considered: rough estimates²² put the upper limit at about 1,300 kb, and we are still finding strong similarity at 3,000 kb (assuming the mean distance in the 1–2 cM range is 1.5 cM). There is considerable variation in the relationship between the genetic and physical maps, however, so absolute conclusions cannot be reached at present. It remains, therefore, to be more fully resolved whether protein rate bands, isochores and G bands are at all inter-related. □

Methods

We compiled a data set of mouse and rat orthologues from scrutiny of entries in HOVERGEN²⁶. Genes were accepted as orthologues if, and only if, the mouse rat sequences had no other non-rodent sequence between them and at least one non-rodent sequence appeared as a sister group. This resulted in a data set of in excess of 500 gene pairs.

Each of the mouse genes was then inspected at LocusLink (www.ncbi.nlm.nih.gov/LocusLink/), by using its accession number, to establish mouse chromosomal location. These chromosomal locations are the same as those described at Mouse Genome Informatics. Those without location specified to the centiMorgan or on the X chromosome were eliminated. (X-linked genes have unusually low synonymous rates of evolution²⁷). This resulted in a data set of 456 genes. Pairwise Blast searching was used to eliminate tandem duplicates from the data set. Any similarity between linked genes led to the elimination of one of the two. This resulted in a data set of 433 autosomal genes; of these, 266 had at least one neighbour within 1 cM.

We used GENETRANS to automatically extract complete coding sequences. DNA alignments were carried out by PILEUP using the default settings. The alignments were checked by eye and modified where necessary. Two genes could not be aligned adequately and were excluded. We estimated substitution rates using a described method²⁸ with modifications²⁹, and applying Kimura's two-parameter method to correct for multiple hits. For each orthologous mouse-rat gene, we therefore obtained values for the rate, per site, for both non-synonymous (K_a) and synonymous (K_s) substitutions. We also calculated the rate of protein evolution controlling for the underlying mutation rate (K_d/K_s).

In the comparison of linked genes, for most loci there was only one neighbour within 1 cM. When there was more than one, the data from any given gene was never used more than twice. In the randomized sets, if a K value was used twice in the pairwise tests, that value was also used twice in the random set. This conserves the number of pairwise comparisons and is also conservative as it permits, in the random set, a difference of zero. The programs for calculating GC4, removal of CpG→TpG mutations, and removal of doublet mutations are in Tcl script.

Received 22 February; accepted 2 June 2000.

- Wolfe, K. H. & Sharp, P. M. Mammalian gene evolution: nucleotide sequence divergence between mouse and rat. *J. Mol. Evol.* **37**, 441–456 (1993).
- Wilson, A. C., Carlson, S. S. & White, T. J. Biochemical evolution. *Ann. Rev. Biochem.* **46**, 573–639 (1977).
- Hurst, L. D. & Smith, N. G. C. Do essential genes evolve slowly? *Curr. Biol.* **9**, 747–750 (1999).
- Graur, D. Amino-acid composition and the evolutionary rates of protein-coding genes. *J. Mol. Evol.* **22**, 53–62 (1985).
- Tourasse, N. J. & Li, W. -H. Selective constraints, amino acid composition, and the rate of protein evolution. *Mol. Biol. Evol.* **17**, 656–664 (2000).
- Kuma, K., Iwabe, N. & Miyata, T. Functional constraints against variations on molecules from the tissue-level—slowly evolving brain-specific genes demonstrated by protein-kinase and immunoglobulin supergene families. *Mol. Biol. Evol.* **12**, 123–130 (1995).
- Hughes, A. L. Circumsporozoite protein genes of malaria parasites (*Plasmodium* spp.): evidence for positive selection on immunogenic regions. *Genetics* **127**, 345–353 (1991).
- Tsaur, S. C. & Wu, C. I. Positive selection and the molecular evolution of a gene of male reproduction, *Acp26Aa* of *Drosophila*. *Mol. Biol. Evol.* **14**, 544–549 (1997).
- Hurst, L. D. & McVean, G. T. Do we understand the evolution of genomic imprinting? *Curr. Opin. Genet. Dev.* **8**, 701–708 (1998).
- Wyckoff, G. J., Wang, W. & Wu, C. I. Rapid evolution of male reproductive genes in the descent of man. *Nature* **403**, 304–309 (2000).
- Makalowski, W. & Boguski, M. S. Synonymous and nonsynonymous substitution distances are correlated in mouse and rat genes. *J. Mol. Evol.* **47**, 119–121 (1998).
- Smith, N. G. C. & Hurst, L. D. The effect of tandem substitutions on the correlation between synonymous and nonsynonymous rates in rodents. *Genetics* **153**, 1395–1402 (1999).
- Matassi, G., Sharp, P. M. & Gautier, C. Chromosomal location effects on gene sequence evolution in mammals. *Curr. Biol.* **9**, 786–791 (1999).

14. Silver, L. M. *Mouse Genetics* (Oxford Univ. Press, Oxford, 1995).
15. Casane, D., Boissinot, S., Chang, B. H. J., Shimmin, L. C. & Li, W. H. Mutation pattern variation among regions of the primate genome. *J. Mol. Evol.* **45**, 216–226 (1997).
16. Averof, M., Rokas, A., Wolfe, K. & Sharp, P. M. Evidence for a high frequency of double-nucleotide substitutions. *Science* **287**, 1283–1286 (2000).
17. Caccio, S. *et al.* Methylation patterns in the isochores of vertebrate genomes. *Gene* **205**, 119–124 (1997).
18. Cargill, M. *et al.* Characterization of single-nucleotide polymorphisms in coding regions of human genes. *Nature Genet.* **22**, 231–238 (1999).
19. Nordborg, M., Charlesworth, B. & Charlesworth, D. The effect of recombination on background selection. *Genet. Res.* **67**, 159–174 (1996).
20. Nachman, M. W. & Churchill, G. A. Heterogeneity in rates of recombination across the mouse genome. *Genetics* **142**, 537–548 (1996).
21. Eyre-Walker, A. Recombination and mammalian genome evolution. *Proc. R. Soc. London Ser. B* **252**, 237–243 (1993).
22. Bernardi, G. The human genome: organization and evolutionary history. *Annu. Rev. Genet.* **29**, 445–476 (1995).
23. Drouin, R., Holmquist, J. P. & Richer, C. L. High-resolution replication bands compared with morphologic G- and R-bands. *Adv. Hum. Genet.* **22**, 47–115 (1994).
24. Bernardi, G. *et al.* The mosaic genome of warm-blooded vertebrates. *Science* **228**, 953–958 (1985).
25. Galtier, N. & Mouchiroud, D. Isochore evolution in mammals: A human-like ancestral structure. *Genetics* **150**, 1577–1584 (1998).
26. Duret, L., Mouchiroud, D. & Gouy, M. HOVERGEN—a database of homologous vertebrate genes. *Nucleic Acid Res.* **22**, 2360–2365 (1994).
27. Smith, N. G. C. & Hurst, L. D. The causes of synonymous rate variation in the rodent genome: Can substitution rates be used to estimate the sex bias in mutation rate? *Genetics* **152**, 661–673 (1999).
28. Li, W.-H. Unbiased estimation of the rates of synonymous and nonsynonymous substitution. *J. Mol. Evol.* **36**, 96–99 (1993).
29. Pamilo, P. & Bianchi, N. O. Evolution of the Zfx and Zfy genes: rates and interdependence between the genes. *Mol. Biol. Evol.* **10**, 271–281 (1993).

Supplementary information is available on Nature's World-Wide Web site (<http://www.nature.com>) or as paper copy from the London editorial office of Nature.

Acknowledgements

We thank W.-H. Li for comments on an earlier version of the manuscript. L.D.H. is funded by The Royal Society.

Correspondence and requests for materials should be addressed to L.D.H. (e-mail: l.d.hurst@bath.ac.uk).

Metapopulation dynamics of bubonic plague

M. J. Keeling* & C. A. Gilligan†

* Department of Zoology, University of Cambridge, Downing Street, Cambridge CB2 3EJ, UK

† Department of Plant Sciences, University of Cambridge, Downing Street, Cambridge CB2 3EA, UK

Bubonic plague is widely regarded as a disease of mainly historical importance; however, with increasing reports of incidence^{1–3} and the discovery of antibiotic-resistant strains of the plague bacterium *Yersinia pestis*⁴, it is re-emerging as a significant health concern^{5,6}. Here we bypass the conventional human-disease models, and propose that bubonic plague is driven by the dynamics of the disease in the rat population. Using a stochastic, spatial metapopulation model, we show that bubonic plague can persist in relatively small rodent populations from which occasional human epidemics arise, without the need for external imports. This explains why historically the plague persisted despite long disease-free periods, and how the disease re-occurred in cities with tight quarantine control. In a contemporary setting, we show that human vaccination cannot eradicate the plague, and that culling of rats may prevent or exacerbate human epidemics, depending on the timing of the cull. The existence of plague reservoirs in wild rodent populations has important public-health implications for the transmission to urban rats and the subsequent risk of human outbreaks.

Large-scale human epidemics of bubonic plague have been recorded throughout history, from Roman times to the pandemic in the early 1900s. This disease has had a major social and demographic effect^{7–10}; its arrival in Europe in 1348 led to the death of around one-third of the human population, and even today bubonic plague kills people in many areas of the world^{1–3}. Historical data, from a variety of locations, show occasional large outbreaks of plague separated by long disease-free periods, and yet the disease clearly persists^{7–9}. Understanding persistence is a common problem in general epidemic modelling^{11–13}, and for bubonic plague it is a central historical question. Previous models of bubonic plague have been highly anthropocentric, modelling the disease as if it were transmitted solely within human populations^{10,14,15}. But consideration of the biology shows that bubonic plague is primarily a disease of rodents that is spread by fleas and only occasionally infects humans; such a disease is termed a zoonosis. From this perspective, we formulate an epizootic (animal-based disease) model for the rat and flea populations, and by coupling this with a standard epidemic (human disease) model, we identify epidemic patterns and the circumstances in which the disease causes a large number of human cases.

The life cycle of the plague can be partitioned into four stages. (1) Fleas feeding on an infected rat ingest the bacteria causing bubonic plague, and soon become infectious. (2) When an infected rat dies, its fleas leave to search for a new host. (3) The fleas usually find other rats, infect them, and so spread the disease through the rodent community. (4) Only when the density of rats is low are the fleas

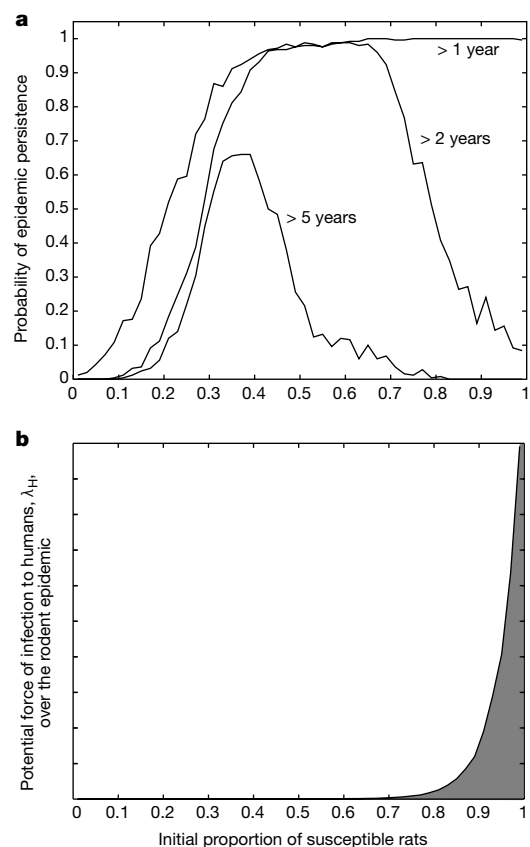


Figure 1 Results from 250 simulations of the stochastic epizootic model of bubonic plague with a rat population of 2,500. **a**, The probability that an infectious import generates an epidemic/endemic that lasts for more than 1, 2 or 5 years in the rat population. If the disease persists for more than 5 years it is likely to be in the endemic state. **b**, The potential force of infection for humans over the entire outbreak in rats ($\int_0^\infty \lambda_{H_i}(t) dt$), measured as the total number of infectious fleas that fail to find a suitable rodent host and may therefore bite and infect humans.

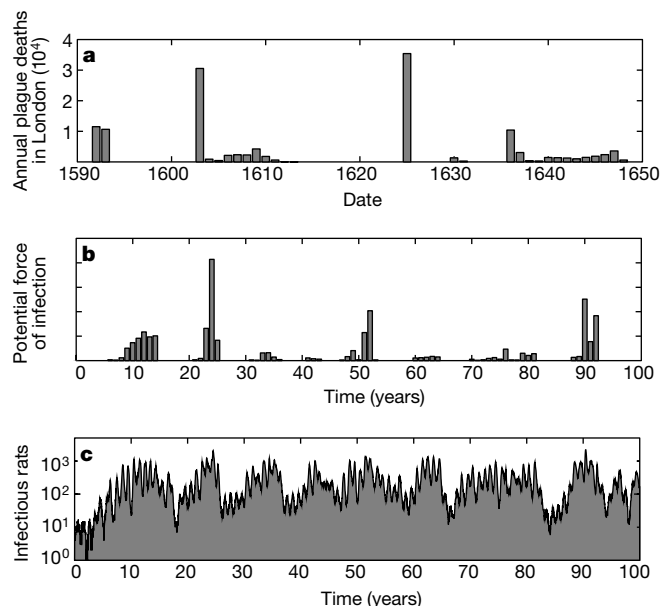


Figure 2 Comparison between historical data and the results from the metapopulation model. **a**, The annual deaths attributed to bubonic plague in London (1592–1661)²⁰. **b** and **c**, Results from a stochastic simulation of 100 years, with a rodent population of 60,000 distributed over 25 subpopulations (see Methods). **b**, Potential force of infection for humans showing erratic epidemics; for comparison with the historical data, yearly aggregates are shown. **c**, Number of infectious rats from the same simulation, showing how plague can persist in an isolated but spatially heterogeneous metapopulation.

forced to feed on alternative hosts such as humans, and a human epidemic occurs. (Although the pneumonic form of the disease can be passed directly from human to human, historical evidence suggests that this is rare^{8,16}, and is only triggered as a consequence of many rat-related cases.) This informal set of rules can be made precise via a set of differential equations, given in the Methods. We identify the potential force of infection for humans, λ_H , which allows us to assess the expected number of human cases. Owing to the slow localized spread of plague in rodents^{8,9}, the dynamics and persistence of the disease in both humans and rats is appropriately considered using a spatially structured version of this model with stochastic dynamics.

Using a metapopulation framework^{17,18}, the community is divided into a set of subpopulations, each of which has its own stochastic dynamics¹⁹ (see Methods). Two phenomena at the local subpopulation scale control the global dynamics of the model (Fig. 1). When the proportion of susceptible rats is low (between 25% and 50%), an import of bubonic plague can often lead to highly persistent endemic behaviour in rats and a very low force of infection to humans. In contrast, when the proportion of susceptible rats is high (greater than 80%) the arrival of infection is followed by a short-lived rat epizootic and a large force of infection to humans (Fig. 1). The global persistence of the disease within the metapopulation depends on a small number of rat subpopulations which are in the endemic state and do not cause human cases; these endemic populations occasionally trigger large but short-lived outbreaks in neighbouring rat subpopulations which in turn lead to major human epidemics.

The potential force of infection for humans, λ_H , from the stochastic metapopulation model can be compared with historical data for mortality in London²⁰ (Fig. 2a and b). Although the predicted number of human cases is erratic with long inter-epidemic periods, plague persists in the rat population (Fig. 2c). In fact, relatively small rodent metapopulations (50,000 individuals) allow the disease to persist globally for many years, although localized extinctions are common. This is attributable to the spatial

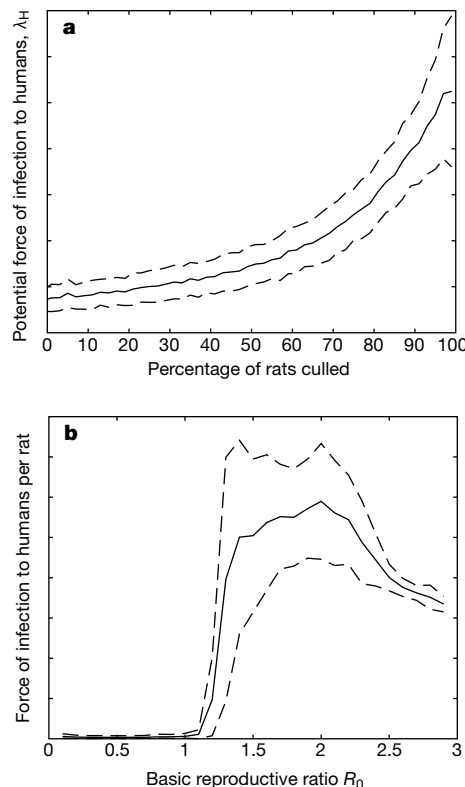


Figure 3 Results from the stochastic model for a single rat subpopulation. Solid lines, means; dashed lines, 95% confidence intervals. **a**, The effect of culling the rats once human cases have been detected. **b**, The potential force of infection to humans per rat over the entire outbreak shows clear threshold behaviour. The basic reproductive ratio R_0 is defined as $R_0 = \beta_R K_F / d_F [1 - \exp(-aK_R)]$. Different values of R_0 were obtained by varying the flea-searching efficiency, **a**, and the rat carrying capacity, K_R .

heterogeneity that develops²¹ and the nonlinear density-dependent behaviour of the disease in rats. The persistence of the disease within the rodent population presents a new perspective on the historical data, as a sudden outbreak of bubonic plague in humans does not have to coincide with fresh imports of the disease from an external source. This resolves a controversy in the historical records for some European cities during the fifteenth and sixteenth centuries, when despite very tight quarantine controls frequent human epidemics still occurred^{8,22}.

Contemporary control or eradication of the plague can be considered using two main approaches, vaccination or culling. However, both methods have associated complications. Because bubonic plague is a zoonosis, primarily passed from rats (via fleas) to humans, there is no ‘herd immunity’ effect^{23,24} associated with vaccinating humans. For most contagious diseases, once the vaccination coverage has exceeded the vaccination threshold (70–80% for smallpox, 90–95% for measles and 99% for malaria) the disease cannot persist in the population²³. For bubonic plague, only those individuals vaccinated are protected, and vaccination of humans cannot provide a means of eradication because the disease is driven by the rat population. An alternative approach is to reduce the rodent population; but our model predicts that there may be difficulties with the timing of any cull. If the rats are kept at a permanently low level, then the risk of a large outbreak in rodents—and therefore the risk of human cases—is reduced. But if the cull is only brought into effect after the first human cases have been reported, then this action can create a far larger force of infection for humans (Fig. 3a). By the time human cases arise, there is already a large reservoir of infection in the rodent population; a cull therefore releases many infected fleas, which may bite humans in

Table 1 Parameter values

Parameter	Value	Meaning	Reference
r_R	5	Rat reproductive rate	25
ρ	0.975	Probability of inherited resistance	27
K_R	2,500	Rat carrying capacity	*
d_R	0.2	Natural rat death rate	25
β_R	4.7	Contact rate	28
m_R	20	(Infectious period) ⁻¹	—
g_R	0.02	Probability of recovery	27
a	0.004	Flea searching efficiency	*
r_F	20	Flea reproductive rate	—
d_F	10	Natural flea death rate	29
K_F	6.57	Flea carrying capacity per rat	27

Details of the model are given in Methods.

*The parameters a and K_R together control the basic reproductive ratio, R_0 , of the disease. The effect of changes in these parameters is therefore described more fully in the text and Fig. 3b. The parameters m_R and r_F have been chosen to be within biologically realistic bounds. Extensive simulation over a range of parameter values shows that the qualitative dynamics of the model have little sensitivity to the precise values used³⁰.

the absence of a suitable rodent host. This highlights the importance of considering bubonic plague as an epizootic infection and explicitly modelling the rat dynamics.

Given the difficulties in control, it is important to calculate the conditions under which a large human outbreak is likely. The recent discovery of antibiotic-resistant strains⁴ increases the importance of the prevention of human cases in the absence of a reliable cure. For many cities in both developing and developed countries, rats are still a serious problem²⁵ and in many of these areas, including the United States, southern Africa, southern Asia and South America, there are large reservoirs of plague in wild rural rodents²⁶. Hence there exists the potential for the disease to enter the highly susceptible urban rat population, with clear public-health consequences.

From the metapopulation model, the potential number of human cases associated with an outbreak is found to depend primarily on the basic reproductive ratio of the disease in rats, R_0 (Fig. 3b), with human cases much more likely above some threshold value. Consideration of the disease in other wild rodent species allows us to place bounds on the flea searching efficiency a , which in turn means that the R_0 threshold corresponds to a rat density of about 3,000 km⁻². Many urban populations exceed this critical density, and identifying those communities at high risk is of obvious public-health importance. □

Methods

Deterministic model

The natural history of the epizootic infection leads to five differential equations describing the dynamics of rats and fleas,

$$\frac{dS_R}{dt} = r_R S_R (1 - T_R/K_R) + r_R R_R (1 - p) - d_R S_R - \beta_R S_R F [1 - \exp(-aT_R)]/T_R$$

$$\frac{dI_R}{dt} = \beta_R S_R F [1 - \exp(-aT_R)]/T_R - (d_R + m_R)I_R$$

$$\frac{dR_R}{dt} = r_R R_R (p - T_R/K_R) + m_R g_R I_R - d_R R_R$$

$$\frac{dN}{dt} = r_F N (1 - N/K_F) + d_F F [1 - \exp(-aT_R)]/T_R$$

$$\frac{dF}{dt} = (d_R + m_R [1 - g_R])I_R N - d_F F$$

$$\lambda_H = F \exp(-aT_R)$$

Here S_R is the number of susceptible rats, I_R is the number of infectious rats, R_R is the number of resistant rats and $T_R = S_R + I_R + R_R$ is the total rat population. In the absence of plague, the rat population experiences logistic growth, with slowly waning resistance to the disease at the population level governed by the parameter f . N gives the average number of fleas on a rat (the flea index), and F is the total number of infected fleas currently searching

for a host. N is again modelled by logistic growth, although it is increased whenever a searching flea finds a new host. Given that fleas search randomly for a new rat host in a limited area, the probability of success is given by $1 - \exp(-aT_R)$, where a is a measure of searching efficiency.

λ_H is the potential force of infection to humans, and as such gives a measure of the number of expected human cases. In fact λ_H is calculated as the number of infected fleas that fail to find a rat host, and therefore is proportional to (but greater than) the rate at which human cases occur. This set of differential equations always shows oscillatory convergence to an equilibrium point with a very low λ_H .

Parameter values

See Table 1.

Stochastic metapopulation

Stochastic spatial dynamics are becoming an increasingly common form of epidemiological modelling. We move from a deterministic framework to a stochastic (or Monte Carlo) simulation, by allowing each possible event (for example, birth, death, infection) to occur randomly but at a rate given by the differential equations¹⁹. To introduce spatial structure, the entire rodent and flea population is discretized into 25 smaller subpopulations in a square lattice arrangement. Each subpopulation has its own stochastic dynamics derived from the differential equations, and adjacent subpopulations are coupled by the random movement of rats and fleas at rates $\mu_R = 0.03$ and $\mu_F = 0.008$, respectively. These movement rates are consistent with the observed slow spread of the disease through communities^{8,9}.

Received 5 April; accepted 23 August 2000.

- Barreto, A., Aragon, M. & Epstein, R. Bubonic plague outbreak in Mozambique 1994. *Lancet* **345**, 983–984 (1994).
- Kumar, S. Bubonic plague in Surat. *Lancet* **345**, 714 (1995).
- Boisier, P. *et al.* Urban epidemic of bubonic plague in Majunga, Madagascar: epidemiological aspects. *Trop. Med. Int. Health* **2**, 422–427 (1997).
- Galimand, M. *et al.* Multidrug resistance in *Yersinia pestis* mediated by a transferable plasmid. *New Engl. J. Med.* **337**, 677–680 (1997).
- Pinner, R. W. Addressing the challenges of emerging infectious disease. *Am. J. Med. Sci.* **311**, 3–8 (1996).
- McCormick, J. B. Epidemiology of emerging/re-emerging antimicrobial-resistant bacterial pathogens. *Curr. Opin. Microbiol.* **1**, 125–129 (1998).
- Shrewsbury, J. F. D. *A History of Bubonic Plague in the British Isles* (Cambridge Univ. Press, Cambridge, 1970).
- Appleby, A. B. The disappearance of plague: a continuing puzzle. *Econ. Hist. Rev.* **33**, 161–173 (1980).
- Curson, P. & McCracken, K. *Plague in Sydney. The Anatomy of an Epidemic* (New South Wales Univ. Press, Kensington, 1986).
- Scott, S., Duncan, C. J. & Duncan, S. R. The plague in Penrith, Cumbria, 1597/8: its causes, biology and consequences. *Ann. Hum. Biol.* **23**, 1–21 (1996).
- Bartlett, M. S. Measles periodicity and community size. *J. R. Statist. Soc. A* **120**, 48–70 (1957).
- Keeling, M. J. Modelling the persistence of measles. *Trends Microbiol.* **5**, 513–518 (1997).
- Grenfell, B. T. & Dobson, A. (eds) *Ecology of Infectious Diseases in Natural Populations* (Cambridge Univ. Press, Cambridge, 1995).
- Noble, J. V. Geographic and temporal development of plagues. *Nature* **250**, 726–728 (1974).
- Raggett, G. F. Modelling the Eyam plague. *Inst. Math. Appl.* **18**, 221–226 (1982).
- Busvine, J. R. *Disease Transmission by Insects* (Springer, Berlin, 1993).
- Grenfell, B. & Harwood, J. (Meta)population dynamics of infectious diseases. *Trends Ecol. Evol.* **12**, 395–399 (1997).
- Hanski, I. A. & Gilpin, M. E. (eds) *Metapopulation Biology. Ecology, Genetics and Evolution* (Academic, San Diego, 1997).
- Renshaw, E. *Modelling Biological Populations in Space and Time* (Cambridge Univ. Press, Cambridge, 1991).
- Graunt, J. *Natural and Political Observations Mentioned In A Following Index, and Made Upon the Bills of Mortality* (Roycroft, London, 1662; available at <http://www.ac.wvu.edu/~stephan/Graunt/bills.html>) (1996).
- Bolker, B. M. & Grenfell, B. T. Space, persistence and dynamics of measles epidemics. *Philos. Trans. R. Soc. Lond. B* **348**, 309–320 (1995).
- Slack, P. The disappearance of plague: an alternative view. *Econ. Hist. Rev.* **33**, 469–476 (1980).
- Anderson, R. M. & May, R. M. *Infectious Diseases of Humans* (Oxford Univ. Press, London, 1992).
- Fox, J. P., Elveback, L., Scott, W., Gatewood, L. & Ackerman, E. Herd-immunity—Basic concept and relevance to public-health immunization practices. *Am. J. Epidemiol.* **94**, 179–189 (1971).
- Buckle, A. P. & Smith, R. H. (eds) *Rodent Pests and Their Control* (Oxford Univ. Press, London, 1994).
- Poland, J. D. & Barnes, A. M. in *CRC Handbook Series in Zoonoses* Sect. A, Vol. 1 (ed Steele, J. F.) 515–556 (CRC, Boca Raton) (1979).
- Hirst, L. F. Plague. *Br. Encyclop. Med. Practice* **9**, 675–698 (1938).
- Macchiavello, A. Reservoirs and vectors of plague. *J. Trop. Med. Hyg.* **57**, 1–68 (1954).
- Bacot, A. W. Observations on the length of time that fleas (*Ceratophyllus fasciatus*) carrying *Bacillus pestis* in the alimentary canals are able to survive in the absence of a host and retain the power to re-infect with the plague. *J. Hyg.* **6**, 770–773 (1915).
- Keeling, M. J. & Gilligan, C. A. Bubonic plague: A metapopulation model of a zoonosis. *Proc. R. Soc. Lond. B* **267**, 2219–2230 (2000).

Acknowledgements

This work was funded by the Royal Society (M.J.K. and C.A.G.), BBSRC and the

Leverhulme Trust (C.A.G.), and arose from the King's College Research Centre project on spatially extended dynamics. We thank S. Betts, B. Grenfell, P. Jones, K. MacCracken, D. MacDonald, F. Nichols, J. Oepfen, R. Smith and J. Swinton for their comments on the manuscript and model parametrization.

Correspondence and requests for materials should be addressed to M.J.K. (e-mail matt@zoo.cam.ac.uk).

Involuntary orienting to sound improves visual perception

John J. McDonald, Wolfgang A. Teder-Sälejärvi & Steven A. Hillyard

Department of Neurosciences, School of Medicine, University of California, San Diego, 9500 Gilman Drive, La Jolla, California 92093-0608, USA

To perceive real-world objects and events, we need to integrate several stimulus features belonging to different sensory modalities. Although the neural mechanisms and behavioural consequences of intersensory integration have been extensively studied^{1–4}, the processes that enable us to pay attention to multi-modal objects are still poorly understood. An important question is whether a stimulus in one sensory modality automatically attracts attention to spatially coincident stimuli that appear subsequently in other modalities, thereby enhancing their perceptual salience. The occurrence of an irrelevant sound does facilitate motor responses to a subsequent light appearing nearby^{5–7}. However, because participants in previous studies made speeded responses rather than psychophysical judgements, it remains unclear whether involuntary auditory attention actually affects the perceptibility of visual stimuli as opposed to postperceptual decision and response processes. Here we provide psychophysical evidence that a sudden sound improves the detectability of a subsequent flash appearing at the same location. These data show that the involuntary orienting of attention to sound enhances early perceptual processing of visual stimuli.

The influence of involuntary auditory attention on speeded motor responses to visual stimuli has been examined in experiments wherein an auditory cue, such as a sudden burst of noise, was presented to the left or right side of a participant who looked at a central spot^{5–11}. A visual target was then presented unpredictably either on the same side as the cue or on the opposite side, and participants had to respond as quickly as possible to the target. Studies involving location-based go/no-go judgements⁵ or elevation discriminations^{6,7} have shown that motor responses were typically faster for targets appearing on the same side as the cue than for targets appearing on the opposite side when the delay between the cue and target was brief (< 300 ms between cue and target onsets). On the basis of these findings, some investigators have proposed that involuntary orienting of attention to a sudden sound enhances perceptual processing of subsequent visual stimuli, as if the sensory responses to those stimuli in the brain were being amplified^{12,13}. However, speeded motor responses cannot be taken as definitive evidence for improvements in perceptual processing because response times are also influenced by a number of postperceptual factors. In particular, the appearance of a spatial cue can modify an observer's willingness to respond¹⁴ and can reduce the uncertainty of an observer's decision¹⁵. The spatial relationship between cue and target can also give rise to stimulus–response compatibility effects and surprise-related disruptions that can affect motor reaction times¹⁶.

Here we used signal detection measures to investigate whether involuntary orienting of attention to a sudden sound influences the

perceptual or postperceptual processing of a subsequent visual stimulus appearing nearby. Unlike reaction times, signal detection measures allow for the separation of perceptual and decision-level effects of attention. In the framework of signal detection theory, the d' parameter reflects an observer's ability to discern a sensory event from its background¹⁷. Thus, if the involuntary orienting of attention to a sound's location facilitated visual perceptual processes, d' should be larger for flashes appearing in proximity to a previous sound than for flashes appearing farther away. Conversely, if the involuntary orienting of attention to a sound's location affected only postperceptual decision processes, then observers might simply require less visual information to decide whether a target is present when it appears closer to the sound source. Such a change in decision would be reflected by a reduction of the decision criterion parameter, β .

We tested these predictions in two cross-modal spatial cueing experiments. The methods were adapted from those recently used to study the effects of visuo-spatial attention on visual detection performance^{18,19}. As shown in Fig. 1, a spatially nonpredictive auditory cue appeared to the left or right side of fixation. This was followed after a brief interval by a visual mask either at the same location (valid trials) or at the opposite location (invalid trials). On half of the trials, a faint visual target was presented at the masked location immediately before the onset of the masking stimulus. On the other half of the trials, the target was absent from the display. Target-present and target-absent trials were randomly intermixed. Participants were informed that the auditory stimulus provided no information about the location of the visual stimuli or whether the target would be present or absent. The task was to press a button to indicate that the target was present, and to refrain from pressing the

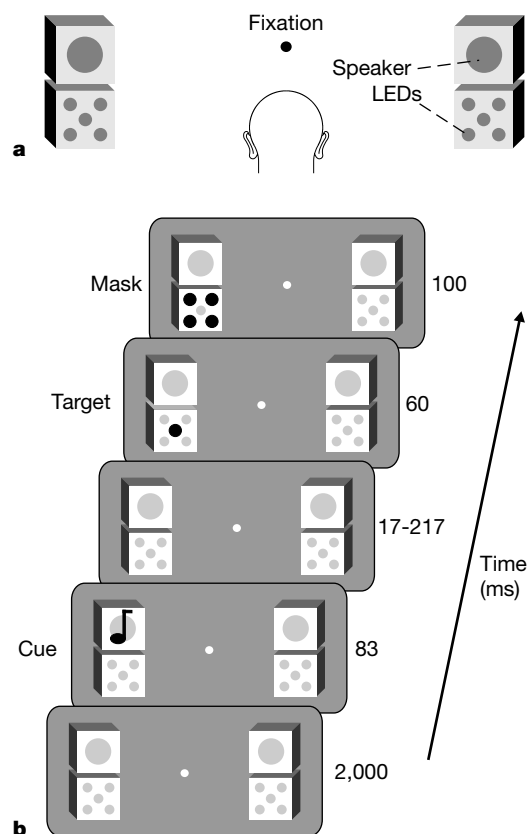


Figure 1 Experimental set-up. **a**, Schematic illustration of the audiovisual apparatus (not drawn to scale). **b**, Illustration of the events occurring on a valid, target-present trial. The cue–target stimulus onset asynchrony is obtained by adding the cue duration and the variable time interval that follows.

same button to indicate that the target was absent. Because only a single mask was presented on any given trial, participants were completely certain about the location at which the target may have occurred. Under such conditions, the observer's report must be based on sensory information arising from that single location, thus eliminating decision-level effects related to reduction of spatial uncertainty^{15,18,19}.

To assess the auditory cue effects on visual detection performance, the mean response times, hit rates, d' values, and β values were subjected to separate analyses of variances with probe location (left versus right) and cue validity (valid versus invalid) treated as within-subject factors. Trials were excluded from the analyses if the participant's eye position deviated from the central fixation spot or if a blink was made. The results from experiment 1 are shown in Fig. 2. On average, participants detected the visual target 20 ms more quickly on valid trials than on invalid trials, $F_{1,14} = 16.4$, $P < 0.001$ (Fig. 2a). Similarly, the mean hit rate (that is, per cent correct responses on target-present trials) was 9% higher on valid trials than on invalid trials, $F_{1,14} = 56.4$, $P < 0.00001$ (Fig. 2b). These data are consistent with recent response time findings⁵⁻⁷. Of primary interest, however, were the results from the signal detection analyses. Perceptual sensitivity (d') averaged 9% higher on valid trials than on invalid trials, $F_{1,14} = 15.6$, $P < 0.001$ (Fig. 2c). This confirms the hypothesis that involuntary orienting of attention to the location of a nonpredictive sound can improve the perceptual quality of a subsequent visual stimulus.

Analysis of the response criterion (β) data showed that involuntary orienting of attention to the location of the auditory cue also influenced the participants' willingness to respond, $F_{1,14} = 10.1$, $P < 0.008$. The decision criterion was on average 29% lower on valid trials than on invalid trials (Fig. 2d). Neither the main effect of probe location nor the probe location $\times$ cue validity interaction approached significance in any of the analyses.

Participants in experiment 1 were instructed to respond as quickly and as accurately as possible. One potential problem with this approach is that the speeded response requirement might reduce the validity of d' as a measure of perceptual sensitivity²⁰. Thus, the effect of cue validity on d' that was observed in experiment

1 might have been partially attributable to the need for rapid responding rather than a genuine effect on perceptual sensitivity. Experiment 2 evaluated this possibility. Experiment 2 was similar to experiment 1, except that response accuracy was emphasized over response speed. Participants were instructed to respond as accurately as possible before the onset of the next warning sound (that is, the cue), which occurred 2–2.5 seconds after the onset of the probe. Upon debriefing, all participants said that they had no difficulties making detection responses within this interval. The intensity of the target was reduced overall to maintain a 70–80% hit rate.

Figure 3 displays the results of experiment 2. Overall, mean response times were longer (552 ms versus 457 ms) and d' s were lower (1.65 versus 1.92) in experiment 2 than in experiment 1. Despite these changes, subjects still detected visual targets more quickly, $F_{1,17} = 21.0$, $P < 0.0002$, and more accurately, $F_{1,17} = 11.3$, $P < 0.0004$, on valid trials than on invalid trials (Fig. 3a and b). In accord with these findings, d' was 8% higher on valid trials than on invalid trials, $F_{1,17} = 6.5$, $P < 0.02$ (Fig. 3c). Thus, experiment 2 confirms that involuntary orienting of attention to a sound improves the perceptual quality of neighbouring visual stimuli. Additionally, the auditory cue had a significant effect on decision processes. As in experiment 1, β was lower on valid trials than on invalid trials, $F_{1,17} = 7.2$, $P < 0.02$ (Fig. 3d). This effect was larger for right visual field stimuli (36%) than for left visual field stimuli (8%), $F_{1,17} = 6.0$, $P < 0.03$. On the basis of these results, we conclude that, under the conditions used here, involuntary attention orienting to a sound influences vision at early perceptual levels as well as at later, decision-related levels.

An irrelevant sound can influence the perception of both concurrent and subsequent visual events. For example, a brief burst of noise can augment the perceived intensity of a concurrent flash of light². In contrast with the present results, however, this intersensory effect did not depend on the relative locations of the auditory and visual stimuli and occurred only when the participants were looking at the flashes. These differences indicate that the effects of a sound on the processing of concurrent and subsequent visual stimuli arise from different neural mechanisms. Neurophysiological studies in animals suggest that the influence of a concurrent sound on

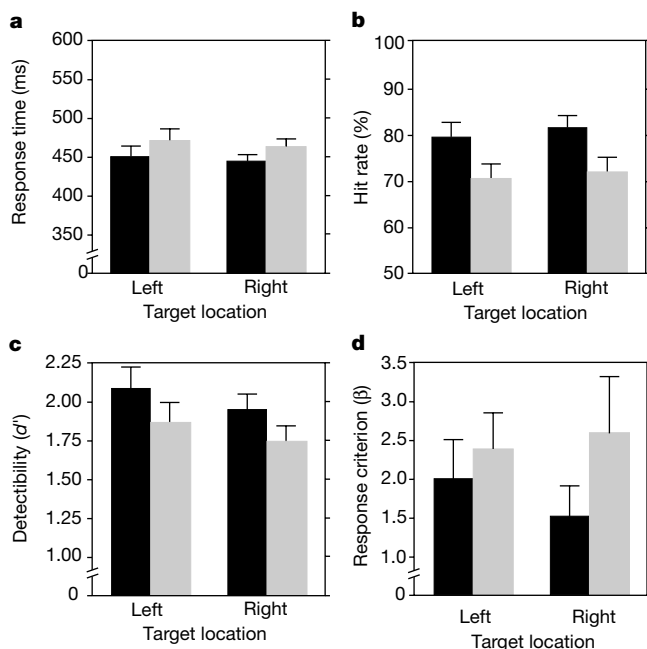


Figure 2 Mean data over all participants from experiment 1. **a**, Response times. **b**, Hit rates. **c**, Perceptual sensitivities (d'). **d**, Response criteria (β). Black bars represent valid conditions and grey bars represent invalid conditions.

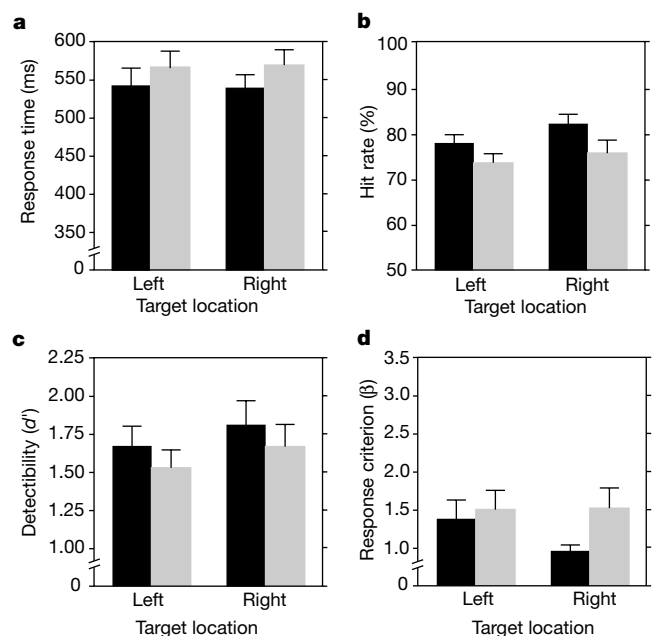


Figure 3 Mean data over all participants from experiment 2. **a**, Response times. **b**, Hit rates. **c**, Perceptual sensitivities (d'). **d**, Response criteria (β). Black bars represent valid conditions and grey bars represent invalid conditions.

visual-evoked brain activity results from the passive integration of auditory and visual inputs by multisensory neurons in the brain^{3,21}. Such integrative effects are minimal when stimuli are separated by 100 ms or more. Because the present design used intervals of 100–300 ms between the auditory cue and visual target, the spatially specific facilitation of visual processing that we observed cannot be explained by passive multisensory integration alone. Instead, the improvement in d' for detection of the visual target can be ascribed to an involuntary orienting of attention to the location of the preceding auditory cue.

It has also been found that a spatially congruent, temporally overlapping sound improved response accuracy (per cent correct) for discriminating the orientation of a 'T' that was surrounded by distractor 'T's but not for discriminating the orientation of a line that was surrounded by distractor lines²². This latter, null effect indicated that visual processing at early, feature-extraction levels may not be altered by concomitant sounds. We found, however, that a preceding sound increased perceptual sensitivity (d') for simple luminance detection under conditions of brief, masked target presentations. This pattern of results is diagnostic of an attention effect upon the early perceptual processing of the spatially congruent visual target^{15,18–20}. This cross-modal facilitation brought about by involuntary spatial attention may be a fundamental operation for enhancing the perceptual salience of natural stimuli in a multisensory world. □

Methods

Participants

Fifteen paid volunteers (8 women) between the ages of 19 and 30 years participated in experiment 1, and eighteen paid volunteers (9 women) between the ages of 19 and 37 participated in experiment 2. Informed, written consent was obtained from each participant.

Stimuli and apparatus

The experiment was conducted in a dimly lit, sound-attenuated chamber. The background sound level within the chamber was 32 dB(A). Participants sat in the chamber and faced a fixation point that was mounted at the centre of a 2.4-m horizontal arc. Loudspeakers mounted on the arc were positioned 38° to the left and right of the central fixation point (see Fig. 1a). A visual display was situated directly below each loudspeaker. Each light display contained four red light-emitting diodes (LEDs) arranged in a 1° square and one green LED positioned at the centre of the square. The distance between the centre of the loudspeaker cone and the centre of the light display was 2°, which is smaller than the minimum audible angle in the far periphery²³. The auditory cue was a 76-dB (A) burst of broadband (500–5,000 Hz) 'pink' noise delivered from one of the two loudspeakers. The visual target was a flash of the green LED on either side of the fixation point. The visual mask was a simultaneous flash of all four red LEDs on either side of the fixation point. The luminance of the mask was 150 cd m⁻², and the luminance of the target varied between 20 and 100 cd m⁻². The luminance of the background was 2 cd m⁻².

Procedure

Each trial began with a 2,000-ms fixation period, during which time the fixation point, loudspeakers and light displays were all visible. Following the fixation period, an 83-ms cue period occurred during which a noise appeared with equal probability from either the left or right speaker. Then, after a random 100 to 300 ms (rectangular distribution), a 60-ms target period occurred. On half of the trials, the target appeared with equal probability on either the left or right side during the target period (target-present trials). On the other half of the trials, no target appeared during the target period (target-absent trials). Immediately following the target period, a mask was presented for 100 ms to either the left or right of fixation. In addition to masking the visual target, the mask indicated the location at which the target may have occurred. Participants were informed that the target and mask never appeared at different locations, and hence their present/absent judgement should be based solely on the masked location. The cue and mask appeared at the same location on 50% of the trials ('valid' trials) and at opposite locations on the other 50% of the trials ('invalid' trials). Thus, the noise cue provided no information about either the location of the mask or whether the target would be present or absent.

Each subject participated in a practice session followed by 15 experimental blocks. The intensity of the target was adjusted during the practice session so that each subject's hit rate was between 70% and 80%. Further adjustments were made between blocks if the subject's hit rate fell outside the 70–80% interval. Each block consisted of 28 valid trials and 28 invalid trials. In experiment 2, there were 14 additional trials per block on which no target or mask occurred. These catch trials were added to reduce anticipatory responses. Participants responded by pressing a button with their left or right thumb on alternating blocks.

Received 8 May; accepted 27 July 2000.

- Bertelson, P. & Aschersleben, G. Automatic visual bias of perceived auditory location. *Psychon. Bull. Rev.* **5**, 482–489 (1998).
- Stein, B. E., London, N., Wilkinson, L. K. & Price, D. D. Enhancement of perceived visual intensity by auditory stimuli: A psychophysical analysis. *J. Cog. Neurosci.* **8**, 497–506 (1996).
- Stein, B. E. & Meredith, M. A. *The Merging of the Senses* (MIT Press, Cambridge, MA, 1993).
- Welch, R. B. & Warren, D. H. in *Handbook of Perception and Human Performance* Vol. 1 (eds Boff, K. R., Kaufman, L. & Thomas, J. P.) **251**, 25–36 (Wiley, New York, 1986).
- McDonald, J. J. & Ward, L. M. Involuntary listening aids seeing: Evidence from human electro-physiology. *Psychol. Sci.* **11**, 167–171 (2000).
- Schmitt, M., Postma, A. & de Haan, E. Interactions between exogenous auditory and visual spatial attention. *Q. J. Exp. Psychol. A* **53**, 105–130 (2000).
- Spence, C. & Driver, J. Audiovisual links in exogenous covert spatial orienting. *Percept. Psychophys.* **59**, 1–22 (1997).
- Mondor, T. A. & Amirault, K. J. Effect of same- and different-modality spatial cues on auditory and visual target identification. *J. Exp. Psychol. Hum. Percept. Perform.* **24**, 745–755 (1998).
- Ward, L. M., McDonald, J. J. & Golestani, N. in *Visual Attention* (ed. Wright, R.) 232–268 (Oxford Univ., New York, 1998).
- Ward, L. M. Supramodal and modality-specific mechanisms for stimulus-driven shifts of auditory and visual attention. *Can. J. Exp. Psychol.* **48**, 242–259 (1994).
- Ward, L. M., McDonald, J. J. & Lin, D. On asymmetries in crossmodal spatial attention orienting. *Percept. Psychophys.* **62**, 1258–1264 (2000).
- Driver, J. & Spence, C. Attention and the crossmodal construction of space. *Trends Cog. Sci.* **2**, 254–262 (1998).
- Driver, J. & Spence, C. Crossmodal attention. *Curr. Opin. Neurobiol.* **8**, 245–253 (1998).
- Sperling, G. & Doshier, B. A. in *Handbook of Perception and Performance* Vol. 1 (eds Boff, K. R., Kaufman, L. & Thomas, J. P.) Ch. 2.1, 2–65 (Wiley, New York, 1986).
- Shiu, L. & Pashler, H. Negligible effect of spatial precuing on identification of single digits. *J. Exp. Psychol. Hum. Percept. Perform.* **20**, 1037–1054 (1994).
- Pashler, H. E. *The Psychology of Attention* (MIT Press, Cambridge, MA, 1998).
- Green, D. M. & Swets, J. A. *Signal Detection Theory and Psychophysics* (Wiley, New York, 1966).
- Luck, S. J., Hillyard, S. A., Mouloua, M. & Hawkins, H. L. Mechanisms of visual-spatial attention: Resource allocation or uncertainty reduction? *J. Exp. Psychol. Hum. Percept. Perform.* **22**, 725–737 (1996).
- Luck, S. J. & Thomas, S. J. What variety of attention is automatically captured by peripheral cues? *Percept. Psychophys.* **61**, 1424–1435 (1999).
- Santee, J. L. & Egeth, H. E. Do reaction time and accuracy measure the same aspects of letter recognition? *J. Exp. Psychol. Hum. Percept. Perform.* **8**, 489–501 (1982).
- Meredith, M. A., Nemitz, J. W. & Stein, B. E. Determinants of multisensory integration in superior colliculus neurons. I. Temporal factors. *J. Neurosci.* **7**, 3215–3229 (1987).
- Dufour, A. Importance of attentional mechanisms in audiovisual links. *Exp. Brain. Res.* **126**, 215–222 (1999).
- Mills, A. W. On the minimum audible angle. *J. Acoust. Soc. Am.* **30**, 237–246 (1958).

Acknowledgements

We thank D. Heraldez for research assistance, M. M. Marlow for computer assistance and L. Anlo-Vento for comments on the manuscript. This work was supported by grants from the National Institute of Health.

Correspondence and requests for materials should be addressed to J.J.M. (e-mail: jmc@sfsu.ca).

Ghrelin induces adiposity in rodents

Matthias Tschöp, David L. Smiley & Mark L. Heiman

Endocrine Research and BioResearch Technologies & Proteins, Lilly Research Laboratories, A Division of Eli Lilly and Company, Lilly Corporate Center, Drop Code 0545, Indianapolis, Indiana 46285, USA

The discovery of the peptide hormone ghrelin, an endogenous ligand for the growth hormone secretagogue (GHS) receptor^{1,2}, yielded the surprising result³ that the principal site of ghrelin synthesis is the stomach and not the hypothalamus. Although ghrelin is likely to regulate pituitary growth hormone (GH) secretion^{3,4} along with GH-releasing hormone and somatostatin, GHS receptors have also been identified on hypothalamic neurons⁵ and in the brainstem⁶. Apart from potential paracrine effects, ghrelin may thus offer an endocrine link between stomach, hypothalamus and pituitary, suggesting an involvement in regulation of energy balance. Here we show that peripheral daily administration of ghrelin caused weight gain by reducing fat utilization in mice and rats. Intracerebroventricular administration of ghrelin

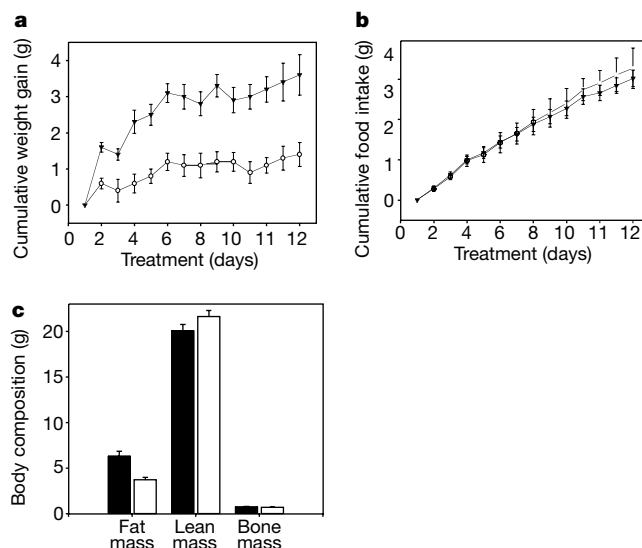


Figure 1 Ghrelin-stimulated adiposity in mice. **a**, Ghrelin induces body weight gain in male wild-type mice ($n = 10$ per group, $P = 0.0001$). Mice treated once daily for two weeks with ghrelin ($2.4 \mu\text{mol kg}^{-1}$, subcutaneously) gained 13.9% of their initial body weight (24.4 ± 1.0 g), while vehicle-injected control animals gained 5.6% of their initial body weight (25.1 ± 1.0 g). **b**, Ghrelin treatment did not change food intake rate in wild-type mice. **c**, Body composition of wild-type mice was measured by DXA after two weeks

of treatment with ghrelin ($2.4 \mu\text{mol kg}^{-1}$, daily subcutaneously) or vehicle ($n = 10$ per group). Mice treated with ghrelin had a greater fat mass (6.34 ± 0.50 g) than vehicle-injected control animals (3.72 ± 0.29 g, $P = 0.002$). Symbols or bars represent the mean $\pm$ the standard error of the mean (s.e.m.). Filled symbols or bars, ghrelin treated; empty symbols or bars, controls.

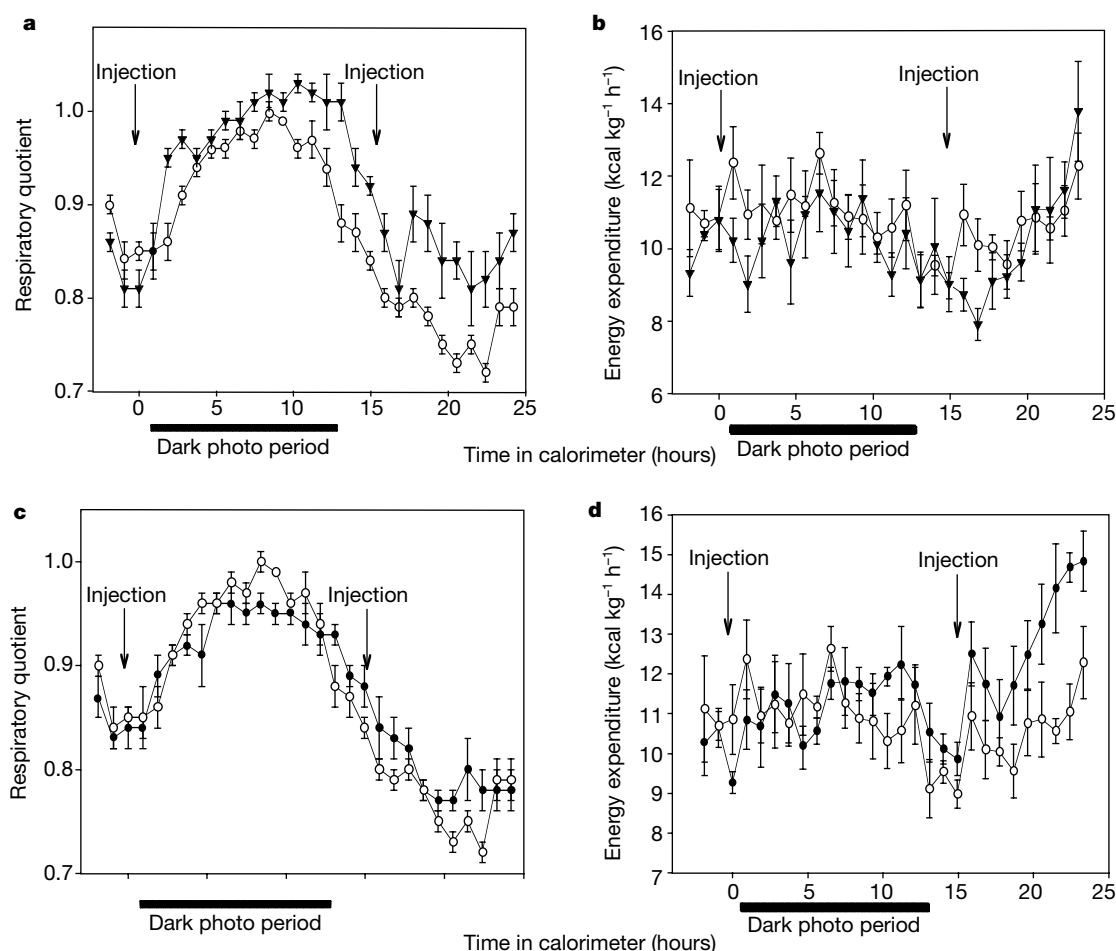


Figure 2 Differential effects of ghrelin and growth hormone on respiratory quotient (RQ) and energy expenditure (EE) in mice. **a**, Ghrelin increased RQ in wild-type mice after each peripheral administration ($P = 0.008$, each group $n = 5$). **b**, Ghrelin did not change total EE after peripheral application in mice. **c**, Growth hormone (8 mg kg^{-1} , subcutaneous) did

not change RQ in wild-type mice (each group $n = 5$, $P = 0.67$) and **d**, did not significantly change total daily EE ($P = 0.18$), but increased EE during the light photoperiod ($P = 0.02$). Symbols represent the mean $\pm$ s.e.m. Filled triangles, ghrelin treated; filled circles, growth hormone treated; open circles, controls.

generated a dose-dependent increase in food intake and body weight. Rat serum ghrelin concentrations were increased by fasting and were reduced by re-feeding or oral glucose administration, but not by water ingestion. We propose that ghrelin, in addition to its role in regulating GH secretion, signals the hypothalamus when an increase in metabolic efficiency is necessary.

We synthesized rat ghrelin and assessed bioactivity by measuring GH release from primary pituitary cells (data not shown). Synthetic rat ghrelin ($2.4 \mu\text{mol kg}^{-1} \text{d}^{-1}$) or vehicle (phosphate buffered saline, PBS) was injected subcutaneously in male wild-type mice (129SV strain) once daily for two weeks. Body weight was significantly increased as compared to vehicle-injected mice (Fig. 1a), although ghrelin did not induce hyperphagia after once daily peripheral administration (Fig. 1b). Analysis of body composition by dual energy X-ray absorptiometry (DXA) after two weeks of ghrelin treatment revealed a significant gain in fat mass, but no change in lean body mass and bone mass (Fig. 1c), bone area ($P = 0.41$) or body length ($P = 0.2$) (data not shown), indicating absence of linear growth stimulation. As the fat gain seemed not to be the result of hyperphagia, we investigated whether ghrelin altered energy expenditure (EE) or decreased the proportion of fuel derived from fat.

Single subcutaneous ghrelin injections ($2.4 \mu\text{mol kg}^{-1}$) during the light photoperiod (resting phase for rodents) and the dark photoperiod (active feeding phase) induced an increase of the respiratory quotient (RQ) ($P = 0.001$, Fig. 2a). Such increased utilization of carbohydrate and reduced utilization of fat to meet energy requirements is congruent with the observed increase in body fat. Neither EE (Fig. 2b) nor locomotor activity (vehicle-injected mice: 70.8 ± 9.6 beam-breaks h^{-1} ; ghrelin-treated mice: 69.2 ± 1.9 beam-breaks h^{-1} ; $n = 5$ per group, $P = 0.9$) were changed

after peripheral application of ghrelin. A selective RQ increase without concomitant increases in carbohydrate intake is unusual and may reflect reduced sympathetic nervous system activity⁸. In addition, direct stimulation of hypothalamic areas can induce a selective change in RQ⁹.

To ascertain the extent that these metabolic data are a consequence of GH release, we injected wild-type mice with GH (8 mg kg^{-1} , subcutaneously). GH did not change RQ during the dark or the light photoperiods (Fig. 2c). In contrast, a single injection of GH significantly increased EE during the light period, reflecting an enhanced energy utilization (Fig. 2d) with no significant alterations in locomotor activity (GH-treated mice: 75.1 ± 12.5 beam-breaks h^{-1} ; vehicle-injected mice: 70.8 ± 9.6 beam-breaks h^{-1} ; $n = 5$ per group, $P = 0.9$). Therefore, induction of a positive energy balance by ghrelin appears unlikely to be caused by its ability to stimulate GH secretion. Daily fat utilization of ghrelin-treated mice ($30.6 \pm 10.5 \text{ kcal kg}^{-1} \text{d}^{-1}$) was significantly less than that measured for GH-treated mice ($68.8 \pm 8.5 \text{ kcal kg}^{-1} \text{d}^{-1}$, $P = 0.024$) or vehicle-injected controls ($65.6 \pm 5.6 \text{ kcal kg}^{-1} \text{d}^{-1}$, $P = 0.038$). Ghrelin-treated mice gained $29.3 \text{ kcal kg}^{-1} \text{d}^{-1}$ of fat (fat intake – fat utilization) while vehicle-injected (control) and GH-treated mice lost $5.7 \text{ kcal kg}^{-1} \text{d}^{-1}$ and $7.2 \text{ kcal kg}^{-1} \text{d}^{-1}$ of fat. Ghrelin-treated mice also tended to increase daily carbohydrate utilization ($130.9 \pm 13.6 \text{ kcal kg}^{-1} \text{d}^{-1}$, $P > 0.17$) as compared to both controls ($103.5 \pm 7.7 \text{ kcal kg}^{-1} \text{d}^{-1}$) and GH-treated mice ($108.4 \pm 8.17 \text{ kcal kg}^{-1} \text{d}^{-1}$).

These data expand the role of ghrelin from a GHS to a hormone that participates in regulating energy balance. Some GHSs are reported to stimulate food intake^{10,11}, and that orexigenic potential may be dependent on both dose and half-life. The described effects may not be mediated by GH, which is known to be lipolytic rather

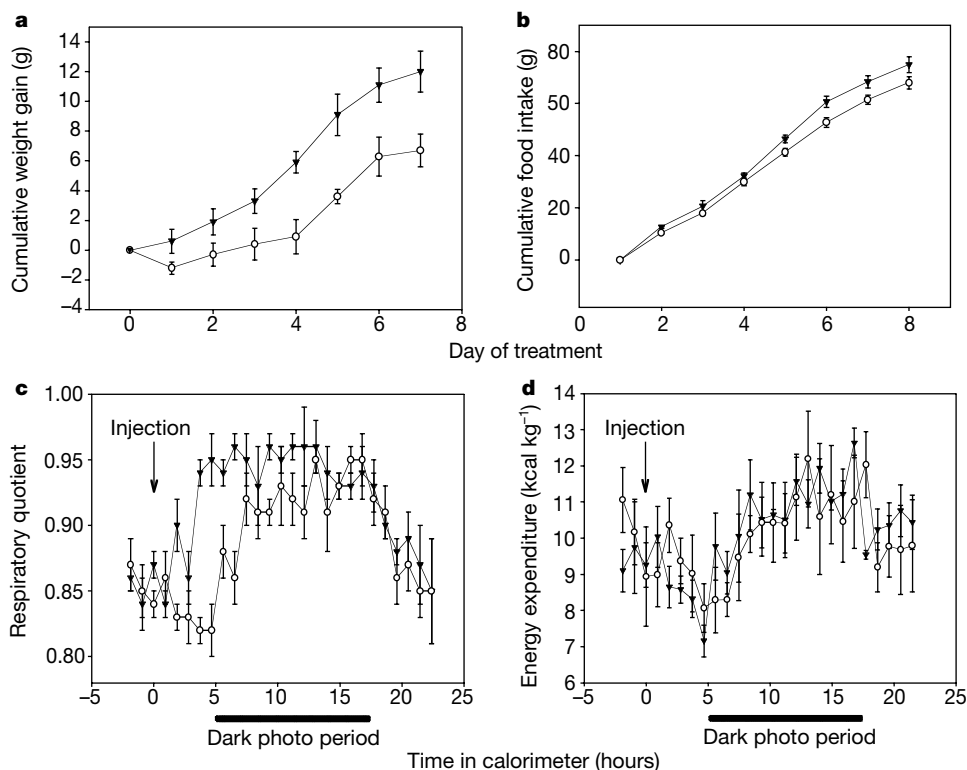


Figure 3 Ghrelin induced a positive energy balance in dwarf rats deficient in growth hormone. **a**, Ghrelin ($4.5 \mu\text{mol kg}^{-1}$ daily, subcutaneous) increased body weight gain in female dwarf rats by 12.1% of their initial weight ($98.9 \pm 1.1 \text{ g}$) when compared to vehicle-injected dwarf rats ($+6.6\%$, initial weight: $101.3 \pm 3.7 \text{ g}$, $P = 0.01$, each group $n = 8$). **b**, Ghrelin tended to increase food intake in dwarf rats ($P = 0.1$). **c**, Ghrelin

($4.5 \mu\text{mol kg}^{-1}$) increased RQ in GH-deficient dwarf rats after a single subcutaneous injection ($n = 5$ per group, $P = 0.002$). **d**, Ghrelin did not change total daily EE in GH-deficient rats, but decreased EE within the first three hours after administration. Symbols represented the mean $\pm$ s.e.m. Filled triangles, ghrelin treated; open circles, controls.

than lipogenic^{12,13}. Further evidence that ghrelin-induced adiposity is independent from ghrelin-mediated GH release was obtained from our studies using GH-deficient dwarf rats¹⁴. An increase in body weight of female dwarfs was observed (Fig. 3a) during one week of ghrelin treatment ($4.5 \mu\text{mol kg}^{-1} \text{d}^{-1}$, subcutaneously). No significant increase in food intake was found, although there was a trend to overeat (Fig. 3b). Like wild-type mice treated with ghrelin, RQ was increased after a single peripheral ghrelin injection (Fig. 3c). Energy expenditure (Fig. 3d) or locomotor activity (data not shown) did not change.

The GHS/ghrelin receptor is present in neuropeptide Y (NPY) neurons^{5,15} and central application of growth hormone-releasing peptide-6 (GHRP-6) induces increased expression of c-Fos protein (an index of cellular activation) in the same neurons^{16,17}. NPY is one of the most effective molecules that induce body weight gain and is likely to mediate the metabolic physiology we observed. To investigate whether ghrelin requires NPY to stimulate fat accretion, we injected NPY-deficient mice¹⁸ with ghrelin ($2.4 \mu\text{mol kg}^{-1} \text{d}^{-1}$) or vehicle for one week. Body weight gain similar to wild-type mice was observed ($P = 0.007$) after one week for ghrelin-treated NPY-deficient mice ($+1.34 \pm 0.30 \text{ g}$) as compared to controls ($-0.14 \pm 0.36 \text{ g}$). A small but significant ($P = 0.032$) increase in food intake was also found in ghrelin-treated NPY-deficient mice (cumulative food intake of ghrelin-treated NPY-deficient mice after one week: $23.7 \pm 1.7 \text{ g}$, controls: $17.6 \pm 2.0 \text{ g}$, $n = 10$ per group). We conclude that the presence of NPY is not obligatory for the ghrelin-stimulated fat accretion.

To determine whether ghrelin-induced adiposity is centrally mediated, we administered ghrelin at very low doses ($\sim 1,000$ -fold

less than peripheral administration) intracerebroventricularly for one week in normal male adult rats ($n = 5$ per group) using mini-osmotic pumps. Continuous central ghrelin administration yielded a dose-dependent, highly significant increase in body weight (Fig. 4a), ($1.2 \text{ nmol kg}^{-1} \text{d}^{-1}$: $+26.7 \pm 10.9 \text{ g week}^{-1}$ ($P = 0.006$); $12 \text{ nmol kg}^{-1} \text{d}^{-1}$: $+37.8 \pm 4.8 \text{ g week}^{-1}$ ($P < 0.001$); controls (water): $+8.6 \pm 5.0 \text{ g week}^{-1}$), enhanced food intake (Fig. 4b), ($1.2 \text{ nmol kg}^{-1} \text{d}^{-1}$: $189.9 \pm 18.5 \text{ g week}^{-1}$ ($P = 0.002$); $12 \text{ nmol kg}^{-1} \text{d}^{-1}$: $204.36 \text{ g week}^{-1}$ ($P < 0.001$); controls: $149.5 \pm 4.4 \text{ g week}^{-1}$), and stimulated mean 24-hour respiratory quotient (Fig. 4c), $1.2 \text{ nmol kg}^{-1} \text{d}^{-1}$: 0.957 ± 0.018 ($P = 0.026$); $12 \text{ nmol kg}^{-1} \text{d}^{-1}$: 0.981 ± 0.007 ($P = 0.02$); controls: 0.907 ± 0.006). No differences in energy expenditure (Fig. 4d), ($P = 0.34$) or locomotor activity ($P > 0.5$) were recorded.

To examine whether endogenous ghrelin is influenced by feeding state, we measured circulating ghrelin levels in normal male Sprague–Dawley rats after 48 hours of fasting, fasting followed by re-feeding for 12 hours, and during *ad libitum* feeding. Serum ghrelin levels were increased in fasted animals ($2.86 \pm 28 \text{ ng ml}^{-1}$, $P < 0.001$) and reduced to levels of *ad libitum* fed rats ($1.26 \pm 14 \text{ ng ml}^{-1}$) after re-feeding ($0.95 \pm 5 \text{ ng ml}^{-1}$, $n = 9$ per group) (Fig. 5a). To investigate whether circulating ghrelin levels are regulated by stomach filling or nutrient intake, we compared oral administration of water (5 ml) with oral gavage of 50% dextrose in water (5 ml) (Fig. 5b and c). Stomach filling with water did not change ghrelin levels, but filling with dextrose significantly reduced serum ghrelin levels ($P = 0.001$).

Thus ghrelin, an endogenous ligand for the GHS-R, induces a positive energy balance in rodents by decreasing fat utilization

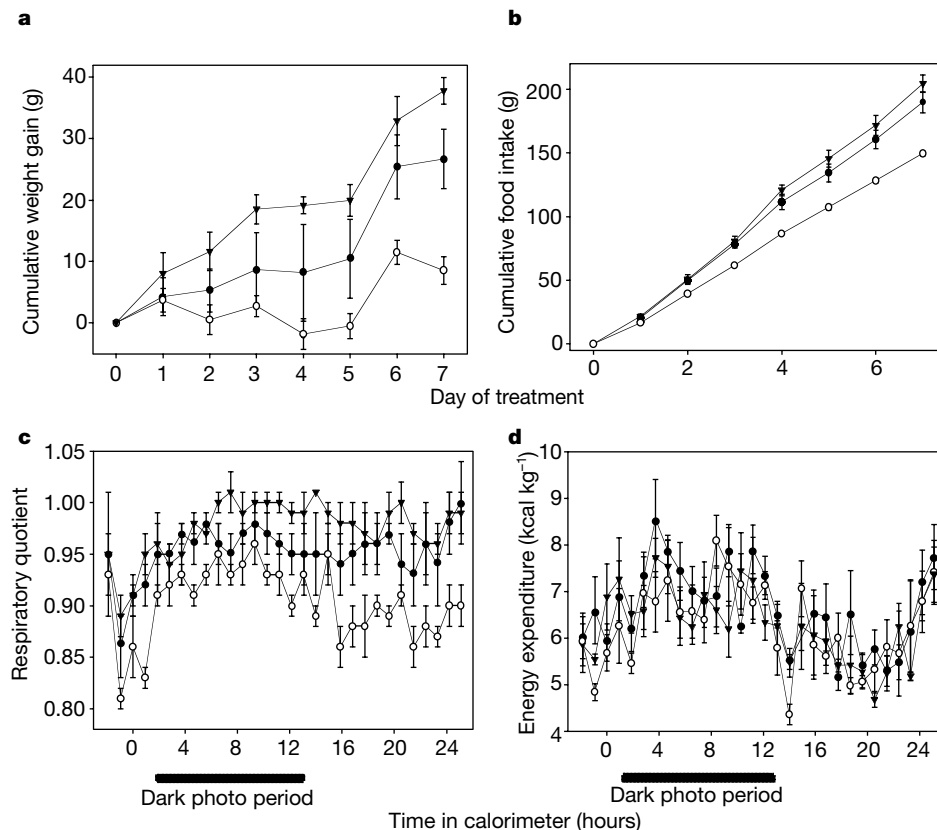


Figure 4 Central effects of ghrelin on energy balance in adult male Sprague–Dawley rats. **a**, Dose-dependent weight gain during continuous intracerebroventricular (ICV) administration of ghrelin ($1.2 \text{ nmol kg}^{-1} \text{d}^{-1}$ and $12 \text{ nmol kg}^{-1} \text{d}^{-1}$) for 7 days ($n = 5$ per group). **b**, Dose-dependent hyperphagia during ICV administration of ghrelin ($1.2 \text{ nmol kg}^{-1} \text{d}^{-1}$ and $12 \text{ nmol kg}^{-1} \text{d}^{-1}$) for 7 days ($n = 5$ per group). **c**, Dose-dependently increased RQ during continuous ICV administration of ghrelin

($1.2 \text{ nmol kg}^{-1} \text{d}^{-1}$ and $12 \text{ nmol kg}^{-1} \text{d}^{-1}$) for 7 days ($n = 5$ per group). **d**, Continuous ICV administration of ghrelin ($1.2 \text{ nmol kg}^{-1} \text{d}^{-1}$ and $12 \text{ nmol kg}^{-1} \text{d}^{-1}$) did not alter energy expenditure of adult rats ($n = 5$ per group). Symbols represented the mean $\pm$ s.e.m. Filled circles, $1.2 \text{ nmol kg}^{-1} \text{d}^{-1}$ ghrelin ICV; filled triangles, $12 \text{ nmol kg}^{-1} \text{d}^{-1}$ ghrelin ICV; open circles, controls.

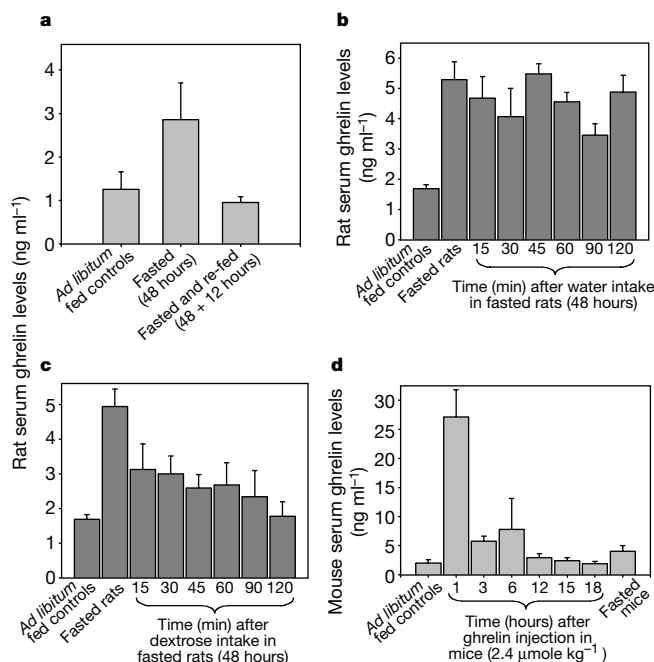


Figure 5 Circulating ghrelin levels respond to feeding state in adult male Sprague-Dawley rats. **a**, Influence of fasting and re-feeding on serum ghrelin levels. Each bar represents mean $\pm$ s.e.m. of 9 animals per group. **b**, Time course of serum ghrelin levels in rats (5 animals per group) after stomach filling with water (5 ml) in comparison to levels of fasted and *ad libitum* fed controls. **c**, Time course of serum ghrelin levels in rats (5

animals per group) after stomach filling with 50% dextrose solution (5 ml) in comparison to levels of fasted and *ad libitum* fed controls. **d**, Time course of serum ghrelin levels in mice (4 animals per group) after a single peripheral (subcutaneous) injection of 2.4 μ mol ghrelin.

without significantly changing energy expenditure or locomotor activity. Serum ghrelin concentration in mice after peripheral (subcutaneous) administration of 2.4 μ mol kg⁻¹ ghrelin was significantly greater than that measured during fasting only within the first hour after injection (Fig. 5d). This indicates that ghrelin has a short half-life. Rapid degradation might offer an explanation for lack of measurable hyperphagia following peripheral administration.

The dramatic diurnal RQ rhythm (Figs 2a and c, 3c and 4c) could be the consequence of a diurnal ghrelin rhythm responding to nutrient absorption postprandially and to an emptying stomach during the rodent's sleep period. Although the peptide is rapidly cleared after peripheral administration (Fig. 5d), it impressively induces hyperphagia after continuous central infusion of doses as low as 1.2 nmol kg⁻¹ day⁻¹ (Fig. 4b), indicating a central mode of action. Ghrelin-induced metabolic changes led to an efficient metabolic state resulting in increased body weight and fat mass. These effects were GH-independent and did not require the presence of NPY. Similar data were observed after administration of growth hormone releasing peptide-2 (GHRP-2), a synthetic agonist of the GHS-R (data will be published elsewhere), indicating a GHS/ghrelin receptor-mediated mechanism. Further, a positive energy balance is necessary to maximize the anabolic actions of GH. As ghrelin stimulates both energy gain and GH secretion, it provides an integrated means to produce an anabolic state. Fasting increases ghrelin serum levels whereas re-feeding blunts this increase, suggesting a role for ghrelin in shifting to a more energy-efficient metabolism during starvation. Sugar intake, but not stomach expansion, decreases circulating ghrelin levels. Although ghrelin is undoubtedly a regulator of GH secretion, our data indicate that this new gastric hormone also signals hypothalamic regulatory centres controlling energy balance. Because ghrelin was detected in the hypothalamus by polymerase chain reaction after reverse transcription of RNA and by immunohistochemistry³, we cannot rule out the possibility that stomach-released ghrelin regulates GH release

whereas central ghrelin influences energy balance. Despite the mechanism, our data indicate that ghrelin may be a new link between the GH/IGF-1 (growth hormone/insulin-like growth factor-1) axis and the neuroendocrine regulation of energy balance. □

Methods

Animals

Wild-type mice (129SV strain) were purchased from Taconic Farms and NPY-knockout mice¹⁸, which were originally provided to us by R. Palmiter, were raised for us at Taconic Farms. Eight-week old dwarf rats¹⁴ were purchased at Harlan, UK. Animals were housed individually in a temperature-controlled environment (25 °C) with a 12-hour light and 12-hour dark (18.00–06.00) photoperiod. All mice had *ad libitum* access to pelleted mouse food and tap water. Calories provided by mouse diet (5008 PMI Nutrition International) were derived from 26.8% protein, 16.7% fat and 56.4% carbohydrate. Calories provided by rat diet (5001 PMI Nutrition International) were derived from 28.0% protein, 12.1% fat and 59.8% carbohydrate. Mice were between 10 and 13 weeks of age and were injected subcutaneously daily between 17.00 and 18.00 with 0.1 ml of PBS containing 0 or 2.4 μ mol kg⁻¹ d⁻¹ ghrelin over 13 days. Food intake and body weights were measured daily at 08.00. Mini-osmotic pumps (Alzet, Alza Pharmaceuticals) were implanted in three groups of five adult male Sprague-Dawley rats for intracerebroventricular (ICV) administration of ghrelin. Pumps were continuously delivering water, 1.2 nmol kg⁻¹ d⁻¹ ghrelin or 12 nmol kg⁻¹ d⁻¹ ghrelin for 7 days at 1.0 μ l h⁻¹. The study protocols used in these experiments were approved by the Animal Care Committee of Eli Lilly and Co. All animal experiments were conducted in accordance with the principles and procedures outlined in the National Institute of Health (NIH) Guide for the Care and Use of Laboratory Animals.

Indirect calorimetry

Twenty-four-hour energy expenditure (EE) and respiratory quotient (RQ) were measured by indirect calorimetry as described¹⁹, using an open circuit calorimetry system (Oxymax, Columbus Instruments Int. Corp.). RQ is the ratio of the volume of CO₂ produced (VCO₂) to the volume of O₂ consumed (VO₂). EE was calculated as the product of calorific value of oxygen (CV) and VO₂ per kilogram of body weight; where CV = 3.815 + 1.232(RQ) (ref. 20). Total calories expended were calculated to determine daily fuel utilization. To calculate the proportion of protein, fat and carbohydrate that is used during that 24-hour period, we used Flatt's proposal²¹ and formulae as well as other derived constants²⁰. Locomotor activity was measured by counting the number of times an animal broke a new light beam during each period of 24 hours in the calorimeter.

Body composition

Body composition was measured *in vivo* on day 14 of the treatment period by dual-energy X-ray absorptiometry (DXA) using a p-DEXA instrument (Norland, USA) as described previously²².

Ghrelin

We synthesized rat ghrelin on an Advanced ChemTech 396 synthesizer and with 9-fluorenylmethoxycarbonyl (Fmoc) amino acids and 50-minute diisopropylcarbodiimide (DIC)/1-hydroxybenzotriazole (HOBt) activated double couplings. Fmoc-SER(Trt) was used in the couplings for Ser3. Following trityl deprotection using 1% TFA / 5% triisopropylsilane in methylene chloride (DCM), the Ser3-hydroxyl was acylated using excess octanoic acid and 1,3-[(dimethylamino)propyl]-3-ethylcarbodiimide hydrochloride salt (EDAC) in the presence of 4-dimethylaminopyridine (DMAP). After removal of the N-terminal Fmoc, a 2-hour cleavage was run using Reagent K. The precipitated peptide was washed with ethyl ether and dried in vacuum. The material was dissolved in aqueous acetic acid and purified over a 2.2 × 25 cm VydacC18 column using a gradient of 15% A to 55% B over 450 min (A = 0.1% TFA, B = 0.1% TFA / 50% CH₃CN). Five-minute fractions were collected while monitoring the ultraviolet at 214 nm (2.0A). The appropriate fractions were combined, frozen and lyophilized. We isolated mono-octanoylated ghrelin and tested for ability to release GH in a primary rat pituitary cell assay. Matrix-assisted laser desorption/ionization time-of-flight mass spectroscopy indicated a mass of 3313.85 g for the purified ghrelin, which was consistent with the theoretical molecular weight.

Bovine growth hormone

Bovine GH was prepared by recombinant DNA technology as described elsewhere²³.

Primary pituitary cell assay

Primary pituitary cells were prepared as described²⁴ and challenged with ghrelin for 15 min at 37 °C. Media was removed and rat GH was measured by radioimmunoassay.

Radioimmunoassays

Rat serum ghrelin levels were measured by a commercially available RIA (Phoenixpeptide) and rat GH level in media from primary pituitary cells was measured by a RIA kit (Amersham).

Statistics

Statistical analysis was performed using one way analysis of variance (ANOVA) with Tukey post-hoc testing using Sigmaplot Software.

Received 7 July; accepted 11 September 2000.

- McKee, K. K. *et al.* Molecular analysis of rat pituitary and hypothalamic growth hormone secretagogue receptors. *Mol. Endocrinol.* **11**, 415–423 (1997).
- Howard, A. D. *et al.* A receptor in pituitary and hypothalamus that functions in growth hormone release. *Science* **273**, 974–977 (1996).
- Kojima, M. *et al.* Ghrelin is a growth-hormone-releasing acylated peptide from stomach. *Nature* **402**, 656–660 (1999).
- Hosoda, H., Kojima, M., Matsuo, H. & Kangawa, K. Purification and characterization of rat des-Gln14-ghrelin, a second endogenous ligand for the growth hormone secretagogue receptor. *J. Biol. Chem.* **275**, 21995–22000 (2000).
- Guan, X. M. *et al.* Distribution of mRNA encoding the growth hormone secretagogue receptor in brain and peripheral tissues. *Brain Res.* **48**, 23–29 (1997).
- Bailey, A. R., Von Engelhardt, N., Leng, G., Smith, R. G. & Dickson, S. L. Growth hormone secretagogue activation of the arcuate nucleus and brainstem occurs via a non-noradrenergic pathway. *J. Neuroendocrinol.* **12**, 191–197 (2000).
- Scheurink, A. J., Leuvenink, H., Benthem, B. & Steffens, A. B. Dexfenfluramine treatment influences plasma catecholamines and energy substrate metabolism in rats. *Physiol. Behav.* **53**, 879–887 (1993).
- Snitker, S., Tataranni, P. A. & Ravussin, E. Respiratory quotient is inversely associated with muscle sympathetic nerve activity. *J. Clin. Endocrinol. Metab.* **83**, 3977–3979 (1998).
- Atrens, D. M., Sinden, J. D., Penicaud, L., Devos, M. & Le Magnen, J. Hypothalamic modulation of energy expenditure. *Physiol. Behav.* **35**, 15–20 (1985).
- Kuriyama, H., Hotta, M., Wakabayashi, I. & Shibasaki, T. A 6-day intracerebroventricular infusion of the growth hormone-releasing peptide KP-102 stimulates food intake in both non-stressed and intermittently-stressed rats. *Neurosci. Lett.* **282**, 109–112 (2000).
- Okada, K. *et al.* Intracerebroventricular administration of the growth hormone-releasing peptide KP-102 increases food intake in free-feeding rats. *Endocrinology* **137**, 5155–5158 (1996).
- Salomon, F., Cuneo, R. C., Hesp, R. & Sonksen, P. H. The effects of treatment with recombinant human growth hormone on body composition and metabolism in adults with growth hormone deficiency. *N. Engl. J. Med.* **321**, 1797–1803 (1989).
- Vernon, R. G. GH inhibition of lipogenesis and stimulation of lipolysis in sheep adipose tissue: involvement of protein serine phosphorylation and dephosphorylation and phospholipase C. *J. Endocrinol.* **150**, 129–140 (1996).
- Charlton, H. M. *et al.* Growth hormone-deficient dwarfism in the rat: a new mutation. *J. Endocrinol.* **119**, 51–58 (1988).
- Willemsen, M. G., Kristensen, P. & Romer, J. Co-localization of growth hormone secretagogue receptor and NPY mRNA in the arcuate nucleus of the rat. *Neuroendocrinology* **70**, 306–316 (1999).
- Bailey, A. R. *et al.* Chronic central infusion of growth hormone secretagogues: effects on fos expression and peptide gene expression in the rat arcuate nucleus. *Neuroendocrinology* **70**, 83–92 (1999).
- Luckman, S. M., Rosenzweig, I. & Dickson, S. L. Activation of arcuate nucleus neurons by systemic administration of leptin and growth hormone-releasing peptide-6 in normal and fasted rats. *Neuroendocrinology* **70**, 93–100 (1999).

- Erickson, J. C., Clegg, K. E. & Palmiter, R. D. Sensitivity to leptin and susceptibility to seizures of mice lacking neuropeptide Y. *Nature* **381**, 415–421 (1996).
- Chen, Y. & Heiman, M. L. Chronic leptin administration promotes lipid utilization until fat mass is greatly reduced and preserves lean mass of normal female rats. *Regul. Pept.* (in the press).
- Elia, M. & Livesey, G. Energy expenditure and fuel selection in biological systems: the theory and practice of calculations based on indirect calorimetry and tracer methods. *World Rev. Nutr. Diet* **70**, 68–131 (1992).
- Flatt, J. P. Assessment of daily and cumulative carbohydrate and fat balances in mice. *J. Nutr. Biochem.* **2**, 193–202 (1991).
- Rose, B. S., Flatt, W. P., Martin, R. J. & Lewis, R. D. Whole body composition of rats determined by dual energy X-ray absorptiometry is correlated with chemical analysis. *J. Nutr.* **128**, 246–250 (1998).
- Hsiung, H. M. & MacKellar, W. C. Expression of bovine growth hormone derivatives in *Escherichia coli* and the use of the derivatives to produce natural sequence growth hormone by cathepsin C cleavage. *Methods Enzymol.* **153**, 390–401 (1987).
- Heiman, M. L., Nekola, M. V., Murphy, W. A., Lance, V. A. & Coy, D. H. An extremely sensitive in vitro model for elucidating structure-activity relationships of growth hormone-releasing factor analogs. *Endocrinology* **116**, 410–415 (1985).

Acknowledgements

We thank J. Caro, G. Cutler Jr, E. Ravussin, R. Al-Awar, C.J. Strasburger and A. Tashjian Jr for critical review, R. Palmiter for providing NPY-deficient mice and L. Craft, J. Baker, J. Bridwell, J. Jacobs, W.T. Johnson, P. Surface, F. Tinsley and T. Butler for technical assistance.

Correspondence and requests for materials should be addressed to M.T. (e-mail: Matthias_Tschoep@Lilly.com).

Repressor activity of Headless/Tcf3 is essential for vertebrate head formation

Cheol-Hee Kim*, Takaya Oda†, Motoyuki Itoh*, Di Jiang*, Kristin Bruk Artinger‡, Settara C. Chandrasekharappa†, Wolfgang Driever§ & Ajay B. Chitnis*

* Laboratory of Molecular Genetics, NICHD and † Genetics and Molecular Biology Branch, NHGRI, NIH, Bethesda, Maryland 20892, USA

‡ Department of Cell Biology, Harvard Medical School, Boston, Massachusetts 02115, USA

§ University of Freiburg, Institute for Zoology, Hauptstrasse 1, D-79104 Freiburg, Germany

The vertebrate organizer can induce a complete body axis when transplanted to the ventral side of a host embryo¹ by virtue of its distinct head and trunk inducing properties. Wingless/Wnt antagonists secreted by the organizer have been identified as head inducers^{2–4}. Their ectopic expression can promote head formation, whereas ectopic activation of Wnt signalling during early gastrulation blocks head formation^{5–7}. These observations suggest that the ability of head inducers to inhibit Wnt signalling during formation of anterior structures is what distinguishes them from trunk inducers that permit the operation of posteriorizing Wnt signals⁸. Here we describe the zebrafish *headless* (*hdl*) mutant and show that its severe head defects are due to a mutation in T-cell factor-3 (Tcf3), a member of the Tcf/Lef family^{9,10}. Loss of Tcf3 function in the *hdl* mutant reveals that *hdl* represses Wnt target genes. We provide genetic evidence that a component of the Wnt signalling pathway is essential in vertebrate head formation and patterning.

The *hdl* mutant was isolated as part of a screen for ethyl nitrosourea (ENU)-induced mutations that disrupt early neurogenesis in zebrafish¹¹. Mutant embryos obtained from *hdl* heterozygous parents, however, display a weak phenotype and are characterized by a slight reduction in eye size. Their weak phenotype allowed a subset of homozygous *hdl* fish to be grown to adulthood. Here, *hdl* mutants refers to maternally and zygotically homozygous mutant

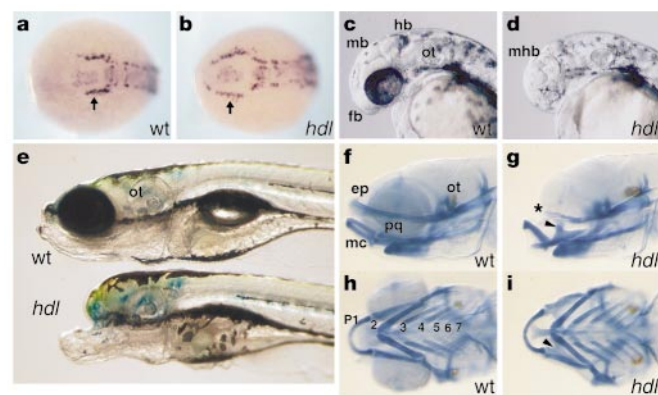


Figure 1 Morphological analysis. **a, b**, *huC* expression at the neural plate stage shows abnormal rostral extension of trigeminal neurons (arrow). Dorsal-anterior view. **c, d**, 34-hour embryos show complete loss of eye, forebrain (fb) and part of the midbrain in *hdl*. mb, midbrain; hb, hindbrain; mhb, midbrain-hindbrain boundary; ot, otic vesicle. **e**, *hdl* larvae at 8 days show a severe head defect. **f–i**, Skeletal preparations of day-5 larvae stained with Alcian blue. In mutants, the ethmoid plate (ep) is reduced (asterisk), whereas ventral pharyngeal arches (p1–p7), including Meckel's cartilage (mc), are less affected. Fused pterygoid process of the palatoquadrate (pq) in *hdl* (arrowhead).

embryos. Our examination of *hdl* mutants with the early neuronal marker, *huC*¹², revealed an aberrant pattern of trigeminal neurons (Fig. 1a, b). The mutation, however, derives its name—*headless*—from the head defect in embryos that is characterized by complete loss of eyes, forebrain and part of the midbrain (Fig. 1c, d, e). Analysis of the head skeleton reveals cranial-specific defects (Fig. 1f, g); the pharyngeal arches appear relatively unaffected (Fig. 1h, i).

Changes in the expression of genes that are early markers for the forebrain and eye are consistent with morphological changes observed in *hdl* mutants. A homeobox gene *anf/Hesx1*, normally expressed in the anterior neural plate (Fig. 2m), is almost absent in mutants (Fig. 2s). Mutational analysis has shown that *Hesx1* is required for normal forebrain development in mice and humans¹³. Transcripts of *six3*, another anterior brain marker¹⁴, are also reduced in the anterior neural plate, but its expression in the underlying prechordal plate is not changed (Fig. 2n, t). This suggests that the *hdl* phenotype is not related to loss of the underlying prechordal plate. Expression of *rx3*, a marker for the presumptive retina and ventral forebrain, is also strongly reduced in mutants (Fig. 2o, u). Mouse embryos carrying a null allele of the *Rx* gene have severe defects in eye and forebrain formation¹⁵. The reduction in expression of these anterior neural-specific genes is accompanied by a rostral expansion of midbrain–hindbrain boundary (MHB) genes such as *pax2* (Fig. 2p, v) and *engrailed2* (*eng2*) (Fig. 2q, w) but not *krox20* (*krox20*), expressed in rhombomeres 3 and 5 (Fig. 2w).

To investigate how early changes in the vertebrate organizer contribute to the genesis of the mutant phenotype, we examined a number of genes expressed in it during early gastrulation. There is little change in the expression of *gooseoid* (data not shown) and *chordin* (Fig. 2a, g) in *hdl* mutants. In contrast, expression of the Wnt antagonist, *dickkopf-1* (*dkk1*)^{3,16}, is increased in the organizer of mutants (Fig. 2b, h). These changes suggest that the *hdl* mutation leads to changes in head induction that start as early as the beginning of gastrulation.

Previous studies have shown that ectopic activation of Wnt signalling after the midblastula transition can mimic some aspects of the *hdl* phenotype^{5–7}, so we examined whether there is ectopic expression of Wnts in the *hdl* mutants. During early gastrulation, *wnt8* is expressed in the ventrolateral margin and excluded from the dorsal margin⁷ (Fig. 2d, e). In *hdl* mutants, *wnt8* expression expands into the dorsal margin (Fig. 2j, k). In addition, rostral expansion of *wnt1* (Fig. 2f, l) and *wnt8b* (data not shown) expression is observed

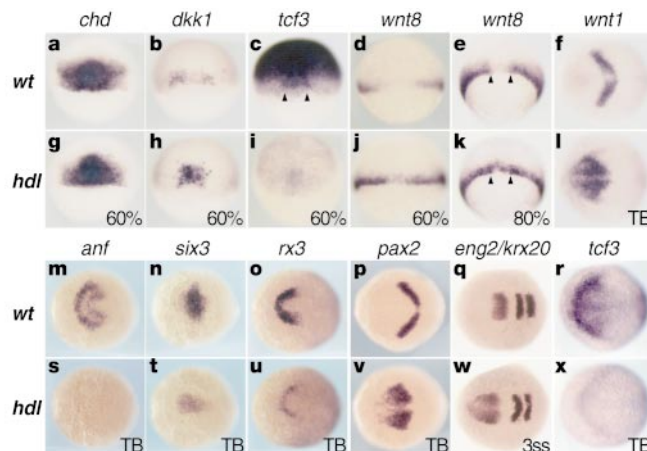


Figure 2 Gene expression analyses. Animal pole is to the top for 60–80% epiboly embryos. Anterior is to the left for tail bud (TB) and 3-somite-stage embryos (3ss). **a, g**, *chd* is relatively unaffected in the organizer. **b, h**, increased *dkk1* expression in *hdl*. **c, i**, Widespread *tcf3* expression, excluded from the ventrolateral margin in the wild type and reduced in *hdl*. Arrowheads indicate the organizer region. **d, e** and **j, k**, *wnt8* is excluded from the dorsal margin (arrowheads) in the wild type but not in *hdl*. **f, l**, Anterior expansion of *wnt1* expression in *hdl*. **m–o** and **s–u**, Reduced expression of anterior neural markers in *hdl*. **p, q** and **v, w**, Anterior expansion of MHB markers. **q, w**, *krox20* is relatively unaffected in *hdl*. **r, x**, Graded anterior neural plate *tcf3* expression, reduced in *hdl*.

in late-gastrula-stage mutant embryos. These data suggest that the loss of anterior neural structures and the rostral expansion of MHB domain in *hdl* mutants may result from ectopic Wnt signalling. To test whether inhibition of Wnt signalling can reduce the *hdl* phenotype, we ectopically expressed the Wnt inhibitors, Cerberus⁴ and Dkk1 (ref. 16). Microinjection of their synthetic RNAs into *hdl* mutant embryos showed that neither of these Wnt inhibitors rescues the mutants (data not shown). This suggests that the normal function of *hdl* is necessary for effective function of these genes.

To clone the *hdl* mutation, we first established its position on the genetic map, using single-strand length polymorphism markers¹⁷. Bulk segregant analysis with centromeric markers localized *hdl* to linkage group 10 (Fig. 3a). Further genetic mapping placed the *hdl* locus between the SSLP markers z6183 and z9701. The proximity of *hdl* to z14193 revealed that the previously isolated *tcf3* gene, a Tcf/Lef family high mobility group (HMG)-box-containing transcription factor¹⁰, mapped close to the *hdl* locus. This gene has maternal expression and is involved in the Wnt pathway, so we investigated *tcf3* as a candidate for the *hdl* gene. We first compared the wild-type sequence of *tcf3* cDNA with that isolated from *hdl* mutants. The wild-type *tcf3* gene encodes a transcription factor with a β -catenin-binding domain, a single HMG DNA-binding domain and a nuclear localization signal (NLS) motif¹⁰. Sequence analysis revealed a seven-base-pair (bp) insertion in the *tcf3* cDNA from *hdl* mutants that causes a frame-shift and is predicted to produce a truncated protein. The truncated Hdl protein retains its β -catenin-binding site but lacks the HMG box as well as the NLS motif (Fig. 3e). Myc-epitope tagged versions of the wild-type and mutant Hdl show that the truncated Hdl proteins are not specifically localized in the nuclei, but the wild-type Tcf3 protein is specifically transported into nuclei (Fig. 3f).

As ENU-based mutagenesis is not likely to cause a 7-bp insertion directly¹⁸, we predicted that the insertion in *tcf3* cDNA from *hdl* mutants was the result of a splicing defect caused by a point mutation. Thus, we isolated the *tcf3* gene from bacterial artificial chromosome (BAC) clones and analysed its genomic structure

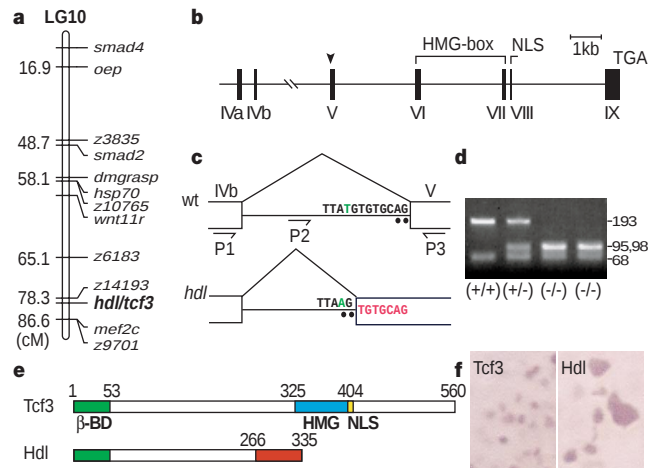


Figure 3 Mapping of the *hdl/tcf3* gene. **a**, Genetic map of linkage group 10 (LG10) showing position of *hdl* in relation to previously mapped genes and SSLP markers. **b**, Structure of the *tcf3* gene. Arrowhead indicates the site of the *hdl* mutation. **c**, T-to-A transversion (green) produces an aberrant splice acceptor site (AG). The 7-bp insertion is indicated in red. **d**, *MseI* digestion of the genomic PCR product (P2–P3), showing a polymorphism. **e**, Structure of Headless (Hdl). The truncated Hdl protein lacks the high mobility group (HMG) box DNA-binding motif as well as the nuclear localization signal (NLS). **f**, Subcellular localization of the wild-type Tcf3 and truncated Hdl proteins is changed from nuclear to cytoplasmic.

(Fig. 3b). Sequence analysis revealed the identical 7-bp sequence in the intron flanking exon V. In addition, we identified a T-to-A transversion in the *hdl* mutant genomic DNA. This point mutation is expected to cause altered splicing in the *hdl* transcripts (Fig. 3c). We used a polymorphism associated with the novel *MseI* site created by the point mutation for fine linkage analysis and showed that no recombinants were seen in 1,018 mutant embryos (Fig. 3d). Sequence analysis also revealed that the structure of the zebrafish *tcf3* gene (Fig. 3b) is similar to the human *Tcf1* gene¹⁹.

The *tcf3* gene is maternally expressed in *Xenopus*⁹ and zebrafish¹⁰, having very wide distribution during the blastula stage. However, despite the broad expression of *tcf3* in the early gastrula, its expression is excluded from the ventrolateral margin (Fig. 2c), where *wnt8* is expressed (Fig. 2d). Strong expression of the *tcf3* transcripts is detectable in the neurectoderm during gastrulation. By the tail bud stage, *tcf3* is expressed in a graded pattern in the anterior neurectoderm, with the highest levels at the anterior edge of the neural plate (Fig. 2r). Tcf3 is absent at the MHB where *wnt1* (Fig. 2f) and *wnt8b* (data not shown) are expressed. These expression patterns of *tcf3* are consistent with its potential early role in head induction and later in patterning the anterior neurectoderm. In *hdl* mutants, *tcf3* expression is strongly reduced both at early blastula and late gastrula stages (Fig. 2i, x).

To confirm that the *hdl* mutation in the *tcf3* gene is responsible for the mutant phenotype, we attempted to rescue the mutant embryos by expressing wild-type *tcf3* RNA. The injection of wild-type *tcf3* RNA (50 pg per embryo) rescued most of the mutants injected (90%; *n* = 50) (Fig. 4a). Ectopic expression of an equivalent amount of the mutant *hdl* RNA could not rescue the headless phenotype (*n* = 50). To test whether the truncated Hdl protein has dominant-negative effects¹⁰, we injected up to a tenfold higher concentration of *hdl* mutant RNA (500 pg per embryo) in both mutant and wild-type embryos. No significant effect was seen in either of these embryos, which suggests that *hdl* is functionally a null mutation.

It is known that Tcf can act as a repressor of Wnt target genes^{20,21} through its association with Groucho^{22,23} and/or CtBP²⁴, or as an activator through its association with β -catenin^{25,26}. Our functional analysis of the *hdl* mutants indicates, however, that the phenotype

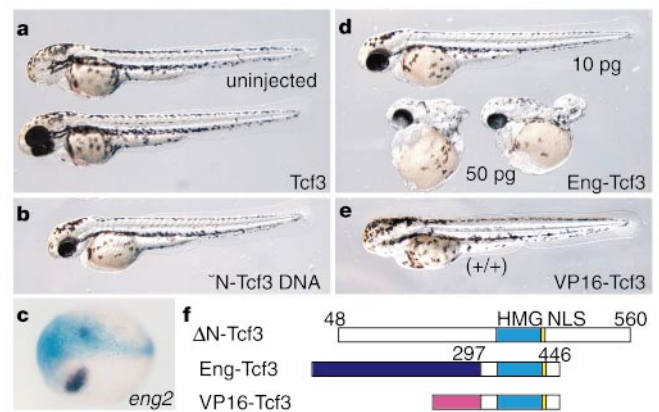


Figure 4 Rescue by Tcf3 expression. **a**, Two-day-old *hdl* mutant injected with wild-type Tcf3 RNA (bottom). Top, uninjected control mutant. **b**, An *hdl* mutant injected with the Δ N-Tcf3 plasmid DNA which is expressed after the MBT. **c**, Δ N-Tcf3 suppresses expression of *eng2*. Blue X-gal staining shows the injected side of embryo. **d**, Rescue of *hdl* by Eng-Tcf3 RNA injection. Higher doses (>50 pg) cause trunk defects. **e**, Opposite effects of VP16-Tcf3 on head formation. A wild-type embryo injected with VP16-Tcf3 RNA, showing a headless phenotype. **f**, Structure of proteins used for injection experiments.

may be associated with the loss of a mechanism that is critical for repressing Wnt signalling, not facilitating it. To test this idea, we examined the effects of Δ N-Tcf3, a deletion of Tcf3 that lacks β -catenin binding (Fig. 4f) but is known to retain its repressor activity⁹. Injection of RNA encoding Δ N-Tcf3 (50 pg per embryo) rescued the *hdl* mutant phenotype (90%; *n* = 50). Embryos injected with plasmid DNA encoding Δ N-Tcf3 also rescued mutants (30%; *n* = 50, Fig. 4b), indicating that zygotic expression of Δ N-Tcf3 is sufficient for rescue. Injection of 50 pg of Δ N-Tcf3 RNA into one blastomere of two-cell-stage embryos suppressed expression of a known target of β -catenin/Tcf signalling²⁷, *eng2*, demonstrating that Δ N-Tcf3 can repress Wnt target genes (Fig. 4c).

To provide further evidence that the activity of Tcf3 as a repressor is critical for head formation, the HMG box DNA-binding domain of Tcf3 was combined with the repressor domain of the Engrailed protein²⁸ (Fig. 4f). Injection of RNA (10 pg per embryo) encoding Eng-Tcf3 rescued the *hdl* mutants (80%; *n* = 40, Fig. 4f). Conversely, we investigated whether a chimaeric transcription activator²⁹ VP16-Tcf3 (Fig. 4f) could rescue *hdl* mutants. None of the mutants were rescued by injection (10 pg per embryo) of VP16-Tcf3 (*n* = 50), although higher doses (>50 pg per embryo) caused severe defects (data not shown). Most wild-type embryos injected with 10 pg of VP16-Tcf3 RNA display a headless phenotype (85%; *n* = 19, Fig. 4e), mimicking the effects of late Wnt activation^{5–7} and the *hdl* mutation.

These results indicate that function of Tcf3 as a repressor may be critical for head formation. Basal repression of Wnt targets by Tcf3 during early gastrulation seems to be essential for repressing the genes responsible for the formation of the MHB domain and for permitting the expression of genes responsible for formation of the forebrain, eyes and the midbrain. Repression of Wnt target genes in the prospective neurectoderm by Tcf3 defines a primary condition for function of head inducers. In the absence of this basal repression, secreted Wnt antagonists are ineffective, because they operate by preventing Wnt signals from de-repressing Wnt target genes in the anterior neurectoderm. The graded pattern of *tcf3* expression at the anterior edge of the neurectoderm and *wnt* genes at the MHB suggest a continuing role for Wnts and their antagonists in patterning the anterior neurectoderm during late gastrulation. □

Methods

Genetic mapping

For mapping, an AB-background heterozygous *hdl* carrier was crossed to two polymorphic strains, WIK and HK. Linkage analysis was performed on embryos obtained by crossing an *hdl* AB homozygous female and an *hdl* AB/HK or *hdl* AB/WIK heterozygous male. Genomic DNA was used for genotyping embryos by polymerase chain reaction (PCR). The genetic distance between SSLP markers and the *hdl* locus was determined by assaying DNA polymorphisms in individual mutant embryos. Zebrafish BAC (192G1 and 103L24, Genome Systems, Inc.) and YAC (76G7, Research Genetics, Inc.) clones were screened by PCR as pools according to the manufacturer's instructions. Restriction enzyme-digested BAC DNA was subcloned into the pBluescript vector and sequenced by standard procedures. For fine linkage analysis, we amplified genomic DNA from individual mutant embryos with specific oligonucleotide primers (P2, 5'-TCAGT-CATTGCGTATCACCTCTGTC-3' and P3, 5'-ACAATGGCGGGGTGAGGGATGC-3') and digested it with the restriction enzyme *MseI*.

Cloning of mutant *hdl* cDNA

RNA was extracted from wild-type embryos or maternally homozygous mutant embryos at the tail bud stage and PCR was carried out after reverse transcription of RNA. Multiple clones were sequenced and compared with wild-type sequences to detect the mutation. To confirm the existence of a 7-bp insertion in the mutant *hdl* cDNA, we amplified a partial cDNA fragment expected to contain the mutated site using the primers P1 (5'-CCTCT TGGATGGTTGGTTCCTCAG-3') and P3. In high-percentage gel electrophoresis, a single predicted 177-bp band, slightly larger than the 170-bp wild-type band, was detected in *hdl* mutant embryos. There were no additional bands detectable in the gel, indicating that the acquired 3'-splice acceptor is preferred for RNA splicing in mutant embryos.

DNA constructs

To generate Myc-tagged Tcf3 (pCS3+MT-Tcf3), first the amino-terminal region, starting from the second amino-acid proline, was generated by PCR using primers flanked by *Bam*HI and *Eco*RI sequences. The PCR product was then subcloned into the *Bgl*II and *Eco*RI site of the pCS3+MT expression vector. Second, the *Bam*HI- and *Spe*I-digested fragment, containing the Myc epitope and the N-terminal β -catenin-binding domain, was subcloned into the CS2+Tcf3a plasmid (a gift from R. Dorsky). pCS3+MT- Δ N-Tcf3 was created by blunt-ligation of the *Nco*I- and *Spe*I-digested pCS3+MT-Tcf3, deleting the first 47 amino acids of β -catenin-binding domain. The pCS2+HA-Eng-Tcf3 construct was generated by fusing the Tcf3 DNA-binding domain (amino acids 297–446) to a fragment encoding amino acids 1–298 of the *Drosophila* Engrailed protein²⁸. pCS2+HA-VP16-Tcf3 was made by fusing the same Tcf3 DNA-binding domain to the activator domain (amino acids 410–490) of the VP16 from the *Herpes simplex* virus²⁹. RNA for injection was prepared from linearized plasmid DNAs using the SP6 mMessage Machine kit (Ambion).

Received 4 July; accepted 4 August 2000.

- Nieto, M. A. Reorganizing the organizer 75 years on. *Cell* **98**, 417–425 (1999).
- Leyns, L., Bouwmeester, T., Kim, S. H., Piccolo, S. & De Robertis, E. M. Frzb-1 is a secreted antagonist of Wnt signaling expressed in the Spemann organizer. *Cell* **88**, 747–756 (1997).
- Glinka, A. et al. Dickkopf-1 is a member of a new family of secreted proteins and functions in head induction. *Nature* **391**, 357–362 (1998).
- Piccolo, S. et al. The head inducer Cerberus is a multifunctional antagonist of Nodal, BMP and Wnt signals. *Nature* **397**, 707–710 (1999).
- Christian, J. L. & Moon, R. T. Interactions between Xwnt-8 and Spemann organizer signaling pathways generate dorsoventral pattern in the embryonic mesoderm of *Xenopus*. *Genes Dev.* **7**, 13–28 (1993).
- Hoppler, S., Brown, J. D. & Moon, R. T. Expression of a dominant-negative Wnt blocks induction of MyoD in *Xenopus* embryos. *Genes Dev.* **10**, 2805–2817 (1996).
- Kelly, G. M., Greenstein, P., Erezylmaz, D. F. & Moon, R. T. Zebrafish *wnt8* and *wnt8b* share a common activity but are involved in distinct developmental pathways. *Development* **121**, 1787–1799 (1995).
- Niehrs, C. Head in the WNT: the molecular nature of Spemann's head organizer. *Trends Genet.* **15**, 314–319 (1999).
- Molenaar, M. et al. XTcf-3 transcription factor mediates β -catenin-induced axis formation in *Xenopus* embryos. *Cell* **86**, 391–399 (1996).
- Pelegrí, F. & Maischein, H. M. Function of zebrafish β -catenin and TCF-3 in dorsoventral patterning. *Mech. Dev.* **77**, 63–74 (1998).
- Artinger, K. B., Chitnis, A. B., Mercola, M. & Driever, W. Zebrafish *narrowminded* suggests a genetic link between formation of neural crest and primary sensory neurons. *Development* **126**, 3969–3979 (1999).
- Kim, C. H. et al. Zebrafish *elav/HuC* homologue as a very early neuronal marker. *Neurosci. Lett.* **216**, 109–112 (1996).
- Dattani, M. T. et al. Mutations in the homeobox gene *HESX1/Hesx1* associated with septo-optic dysplasia in human and mouse. *Nature Genet.* **19**, 125–133 (1998).
- Kobayashi, M., Toyama, R., Takeda, H., Dawid, I. B. & Kawakami, K. Overexpression of the forebrain-specific homeobox gene *six3* induces rostral forebrain enlargement in zebrafish. *Development* **125**, 2973–2982 (1998).
- Mathers, P. H., Grinberg, A., Mahon, K. A. & Jamrich, M. The Rx homeobox gene is essential for vertebrate eye development. *Nature* **387**, 603–607 (1997).
- Hashimoto, H. et al. Zebrafish *Dkl1* functions in forebrain specification and axial mesoderm formation. *Dev. Biol.* **217**, 138–152 (2000).
- Gates, M. A. et al. A genetic linkage map for zebrafish: comparative analysis and localization of genes and expressed sequences. *Genome Res.* **9**, 334–347 (1999).

- Mullins, M. C., Hammerschmidt, M., Haffter, P. & Nusslein-Volhard, C. Large-scale mutagenesis in the zebrafish: in search of genes controlling development in a vertebrate. *Curr. Biol.* **4**, 189–202 (1994).
- van de Wetering, M. et al. The human T cell transcription factor-1 gene. Structure, localization, and promoter characterization. *J. Biol. Chem.* **267**, 8530–8536 (1992).
- Lin, R., Thompson, S. & Priess, J. R. pop-1 encodes an HMG box protein required for the specification of a mesoderm precursor in early *C. elegans* embryos. *Cell* **83**, 599–609 (1995).
- van de Wetering, M. et al. Armadillo coactivates transcription driven by the product of the *Drosophila* segment polarity gene *TCF*. *Cell* **88**, 789–799 (1997).
- Cavallo, R. A. et al. *Drosophila* Tcf and Groucho interact to repress Wingless signalling activity. *Nature* **395**, 604–608 (1998).
- Roose, J. et al. The *Xenopus* Wnt effector XTcf-3 interacts with Groucho-related transcriptional repressors. *Nature* **395**, 608–612 (1998).
- Brannon, M., Brown, J. D., Bates, R., Kimelman, D. & Moon, R. T. XTcfBP is a XTcf-3 co-repressor with roles throughout *Xenopus* development. *Development* **126**, 3159–3170 (1999).
- Clevers, H. & van de Wetering, M. TCF/LEF factor earn their wings. *Trends Genet.* **13**, 485–489 (1997).
- Moon, R. T., Brown, J. D. & Torres, M. WNTs modulate cell fate and behavior during vertebrate development. *Trends Genet.* **13**, 157–162 (1997).
- McGrew, L. L., Takemaru, K., Bates, R. & Moon, R. T. Direct regulation of the *Xenopus* engrailed-2 promoter by the Wnt signaling pathway, and a molecular screen for Wnt-responsive genes, confirm a role for Wnt signaling during neural patterning in *Xenopus*. *Mech. Dev.* **87**, 21–32 (1999).
- Badiani, P., Corbella, P., Kioussis, D., Marvel, J. & Weston, K. Dominant interfering alleles define a role for c-Myb in T-cell development. *Genes Dev.* **8**, 770–782 (1994).
- Trizzenberg, S. J., Kingsbury, R. C. & McKnight, S. L. Functional dissection of VP16, the trans-activator of herpes simplex virus immediate early gene expression. *Genes Dev.* **2**, 718–729 (1988).

Acknowledgements

We thank R. T. Moon, R. Dorsky, T. Hirano, M. Hibi, A. Kawahara, L. Kodjabachian, D. Turner, M. Halpern and M. Kobayashi for constructs; M. Kacergis and G. Palardy for technical assistance; R. Subramanian and members of the Dawid lab for comments on the manuscript. The mutagenesis screen was performed at CVRC/MGH/Harvard Medical School. This work was supported by the NIH.

Correspondence and requests for materials should be addressed to A.B.C. (e-mail: chitnisa@mail.nih.gov).

Development of Th1-type immune responses requires the type I cytokine receptor TCCR

Qi Chen*, Nico Ghilardi*, Hua Wang†, Thad Baker†, Ming-Hong Xie‡, Austin Gurney‡, Iqbal S. Grewal†§ & Frederic J. de Sauvage*§

Departments of *Molecular Oncology, †Immunology and ‡Molecular Biology, Genentech, Inc., 1 DNA Way, South San Francisco, California 94080, USA
§These authors contributed equally to this work.

On antigen challenge, T-helper cells differentiate into two functionally distinct subsets, Th1 and Th2, characterized by the different effector cytokines that they secrete¹. Th1 cells produce interleukin (IL)-2, interferon- γ (IFN- γ) and lymphotoxin- β , which mediate pro-inflammatory functions critical for the development of cell-mediated immune responses, whereas Th2 cells secrete cytokines such as IL-4, IL-5 and IL-10 that enhance humoral immunity^{1,2}. This process of T-helper cell differentiation is tightly regulated by cytokines. Here we report a new member of the type I cytokine receptor family, designated T-cell cytokine receptor (TCCR). When challenged *in vivo* with protein antigen, TCCR-deficient mice had impaired Th1 response as measured by IFN- γ production. TCCR-deficient mice also had increased susceptibility to infection with an intracellular pathogen, *Listeria monocytogenes*. In addition, levels of antigen-specific immunoglobulin- γ 2a, which are dependent on Th1 cells, were markedly reduced in these mice. Our results demonstrate the existence of a new cytokine receptor involved in regulating the adaptive immune response and critical to the generation of a Th1 response.

The expansion and terminal differentiation of T-helper cells from a common precursor³ into distinct populations of Th1 and Th2 effector cells is controlled by several factors, including the cytokine

microenvironment during lymphocyte activation by antigen^{1,2}. Members of the family of four-helix-bundle cytokines⁴ have been found to be critical for this process⁵. IL-4 influences predominantly the development of Th2 cells, whereas IL-12 is a principal factor involved in the differentiation of Th1 cells^{6–9}. Accordingly, mice deficient in IL-4 (ref. 10), IL-4 receptor α -chain¹¹, or the IL-4-specific transcription factor STAT6 (refs 12–14) are defective in Th2 responses, whereas mice deficient in IL-12 (ref. 15), IL-12 receptor (IL-12R) β 1 chain¹⁶ or the IL-12-specific transcription factor STAT4 (ref. 17) have impaired Th1 responses. Additional factors are, however, thought to be involved as none of these mice displays a complete absence of Th1 or Th2 response upon antigen stimulation.

Using a homology screen, we identified TCCR, a member of the type I cytokine receptor family that is most homologous to the IL-12 receptor β 2 chain (26% identity, 37% similarity). From a peripheral blood leukocyte (PBL) library, we isolated a full-length human TCCR complementary DNA that encoded a protein of 636 amino acids with a single transmembrane domain, a WSX signature motif (W-G-x-W-S) in its extracellular domain and a box1 motif in its intracellular region (Fig. 1a). Mouse TCCR isolated from a spleen library revealed 62% identity when compared with the human sequence. Examination of human tissues by northern blot analysis using TCCR as a probe showed the presence of a 3.5-kilobase (kb) transcript in adult lymphoid tissues including PBLs, thymus and spleen (Fig. 1b). Expression was also detected in fetal and adult lung. To determine which cell types express TCCR, we separated murine splenocytes into CD4⁺ T cells, B cells, natural killer cells, macrophages and CD8⁺ T cells, and analysed TCCR expression by real-time polymerase chain reaction (PCR)¹⁸. TCCR expression was detected in all cell types examined, with highest levels in CD4⁺ T cells and natural killer cells (Fig. 1c). TCCR expression was highest in Th0 cells and appeared to be downregulated on differentiation of CD4⁺ T cells in both Th1 and Th2 cells (Fig. 1d).

To determine the role of the TCCR *in vivo*, we generated TCCR-deficient mice by gene targeting. A targeting vector in which exons 2–8 were replaced by a PGK1-*neo* cassette (Fig. 2a) was transfected into embryonic stem cells. Gene targeting was detected in 1 out of 150 embryonic stem (ES) colonies, and 3 clones were selected for micro-injection into blastocysts (Fig. 2b). Founder mice with appropriate germline transmission were interbred to generate homozygous gene-targeted mice (Fig. 2c). Analysis of splenic RNA by northern blot and PCR with reverse transcription (RT-PCR) showed lack of functional TCCR transcript in TCCR-deficient mice, indicating that the TCCR gene had been properly targeted (Fig. 1d; and data not shown). TCCR-deficient mice were viable, fertile and displayed no overt abnormalities. Detailed histological examination did not reveal any obvious defects (data not shown). Flow cytometry analysis of cells obtained from thymus, spleen, lymph nodes and Peyer's patches of several wild-type and knockout mice stained with antibodies to CD3, CD4, CD8, CD25, CD19, B220, CD40, NK1.1, DX5, F4/80, CD14, CD16, MHC II and CD45 did not reveal any gross differences between the two genotypes (data not shown).

As TCCR expression is predominant in lymphoid tissues, we investigated the role of TCCR in the generation of an immune response *in vivo*. We immunized wild-type and TCCR-deficient mice with keyhole limpet haemocyanin (KLH), collected draining lymph nodes 9 d later and assessed them for cytokine production after *in vitro* stimulation with KLH. The ability of TCCR-deficient cells to produce IFN- γ was significantly impaired when challenged with KLH, whereas the production of IL-4 was markedly enhanced (Fig. 3a, b). Production of IL-5 and antigen induced proliferation of TCCR-deficient lymph-node cells that had been primed *in vivo* were normal (Fig. 3c, d). Normal levels of IFN- γ production were measured upon LPS stimulation of TCCR-deficient mice (data not shown), indicating that there seemed to be no intrinsic defects in IFN- γ production in these mice. These results indicate that

TCCR-deficient mice are impaired in their ability to mount a Th1 response. The loss of Th1 response is accompanied by an enhanced Th2 response similar to that observed in mice deficient in Th1 cytokines such as IL-12 (refs 15, 16).

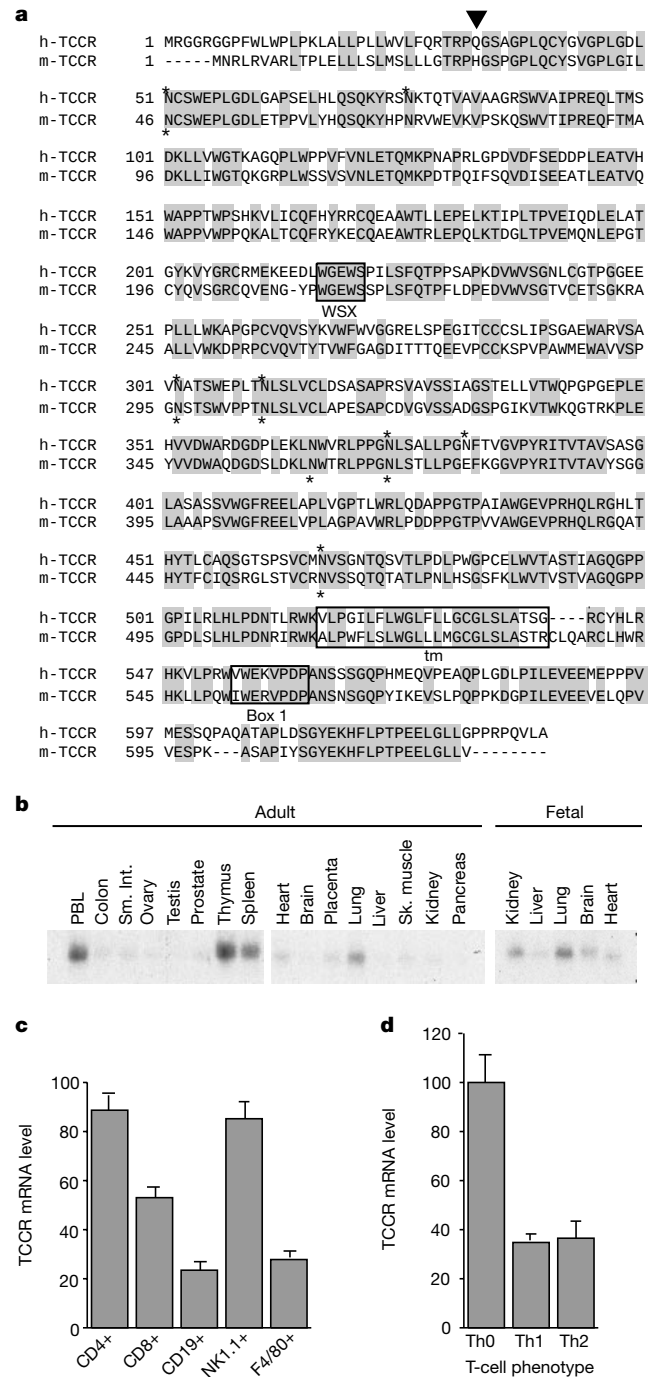


Figure 1 Amino-acid sequence and tissue distribution of TCCR. **a**, Comparison of the human and mouse TCCR sequences. Identical amino acids are shaded and gaps introduced for optimal alignment are indicated by dashes. The predicted signal peptidase cleavage site is indicated by an arrowhead. Potential *N*-glycosylation sites are indicated with an asterisk. The WSX motif, transmembrane domain and box1 motif are boxed. **b**, Northern blot analysis of human TCCR. Human multiple-tissue northern blots (Clontech) were probed with a cDNA fragment corresponding to the extracellular domain of TCCR. Sm. Int., small intestine; Sk., skeletal. **c**, **d**, Expression of TCCR messenger RNA in sorted murine splenocytes (**c**), and in undifferentiated CD4⁺ T cells (Th0) or CD4⁺ T cells differentiated for 3 d into Th1 or Th2 cells (**d**). RNA levels were determined by Taqman real-time quantitative PCR and were normalized to rpl19, a ribosomal housekeeping gene.

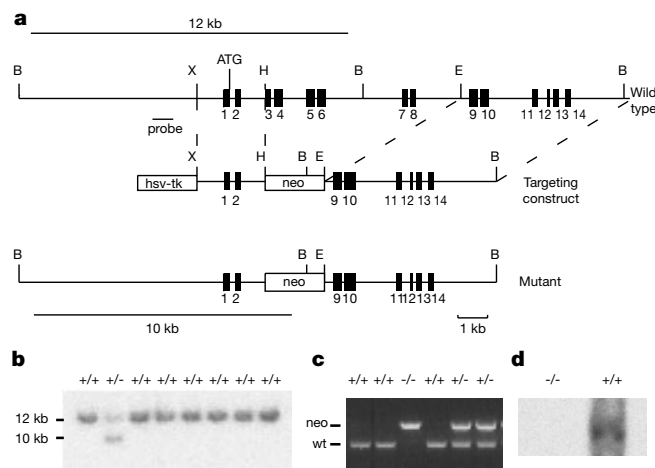


Figure 2 Targeting of the TCCR gene by homologous recombination. **a**, Structure of the TCCR gene and gene targeting vector. A neomycin resistance gene under control of the PGK promoter replaced exons 3–8, deleting most of the TCCR extracellular domain. B, *Bam*H1; E, *Eco*RI; H, *Hind*III; X, *Xba*I. **b**, Targeting of the TCCR gene in ES cells. An oligonucleotide probe was designed from sequences 5' to those present on the targeting vector. After homologous recombination, the mutated allele was detected by the addition of a *Bam*H1 site from the *neo* gene. **c**, PCR analysis of genomic DNA isolated from mouse

tail snips of the offsprings of interbreeding heterozygous mice. The DNA was amplified with a primer set specific for the *Neo* gene (Mutant) and a primer set which amplify a fragment deleted in the targeted allele and therefore specific for the wild-type allele (wt). **d**, Northern blot analysis of TCCR expression in TCCR knockout and wild-type mice. RNA was prepared from the spleen of the indicated mice, separated on a 1% agarose-formaldehyde gel, transferred to a nylon membrane and hybridized with a TCCR DNA probe.

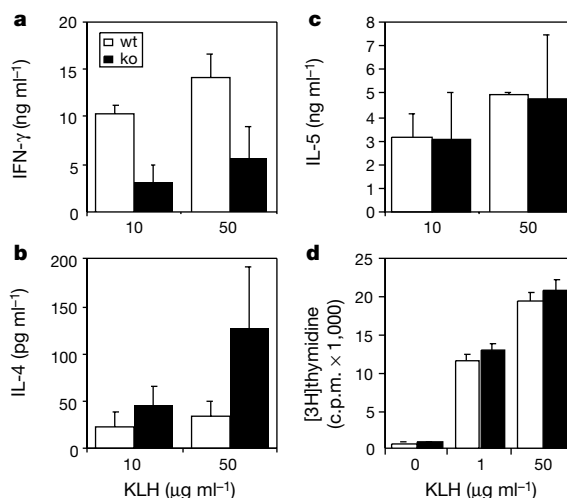


Figure 3 Antigen-induced cytokine production and proliferation by lymph-node cells from TCCR-deficient mice. Wild-type and TCCR-deficient mice were immunized with KLH in CFA. Lymph nodes were collected 9 d later and cultured in the presence of KLH as indicated and analysed for their capacity to produce IFN- γ (**a**), IL-4 (**b**) or IL-5 (**c**), or to

proliferate (**d**). Data are presented as the mean $\pm$ s.d. values that were derived from five animals in each group. $P < 0.004$ (unpaired *t*-test) for IFN- γ levels between wild-type and knockout mice at both KLH concentrations.

In addition to its role in regulating the cellular immune response, IFN- γ is also involved in immunoglobulin isotype regulation. In particular, IFN- γ enhances the production of IgG2a antibodies and, to a lesser extent, of IgG3 antibodies^{19,20}. Consistent with a diminished production of IFN- γ by Th1 cells, TCCR-deficient mice had decreased total serum IgG2a concentrations but normal levels of all other immunoglobulin isotypes when compared with wild-type littermates (Fig. 4a). Furthermore, upon *in vivo* challenge with ovalbumin (OVA), TCCR-deficient mice had severely reduced titres of OVA-specific IgG2a ($\sim 20\%$ of controls; Fig. 4b).

Th1 response is crucial in the defence against intracellular pathogens such as *Listeria monocytogenes*. To further establish the *in vivo* role of TCCR in the control of Th1 response, we infected TCCR-deficient mice and control littermates with a sublethal dose of *L. monocytogenes* (3×10^4 colony forming units (c.f.u.)). Bacterial titres were determined 3 or 9 d after infection and found to be up to

10^6 -fold higher in the livers of TCCR-deficient mice (Fig. 4c).

We next investigated the role of TCCR in mediating the differentiation of a Th1 response *in vitro*. CD4⁺ T cells from wild-type and TCCR-deficient mice were differentiated *in vitro* in the presence of irradiated antigen-presenting cells (APCs) under conditions that favour cell development of either Th1 or Th2. After 4 d in culture, cells were washed and re-stimulated with ConA, and 24 h later supernatants were analysed for the presence of cytokines²¹. When differentiated into Th1 cells, TCCR-deficient lymphocytes produced 80% less IFN- γ than the lymphocytes of wild-type littermates (Fig. 5a). In contrast, TCCR-deficient lymphocytes grown in the presence of IL-4 and anti-IFN- γ antibodies produced slightly more IL-4. Similar results were obtained with CD4⁺ CD45Rb^{high} naive T cells (see Supplementary Information). This effect is intrinsic to the T cells for two reasons: first, similar results were obtained when T cells were differentiated in the presence of APCs

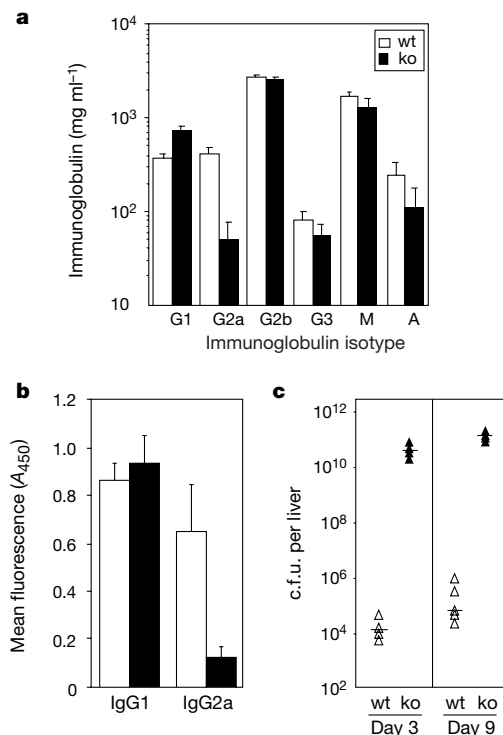


Figure 4 Effect on IgG subclass concentrations and sensitivity to *L. monocytogenes* infection. **a**, Serum was collected from wild-type and TCCR-deficient mice, and total IgG subclass concentrations were determined by ELISA. **b**, OVA-specific IgG1 and IgG2a from OVA/CFA primed mice. Serum was collected from wild-type and TCCR-deficient mice that were immunized with OVA in CFA and levels of IgG1 (1:320,000 dilution) and IgG2a (1:5,000 dilution) were determined by OVA-specific ELISA. **c**, Five TCCR-deficient mice or wild-type littermates were infected subcutaneously with 3×10^4 c.f.u. of *L. monocytogenes*. Three or nine days later, the livers were collected and bacterial titres determined. Data are presented as the mean $\pm$ s.d. values that were derived from five animals in each group. $P < 0.001$ (unpaired *t*-test) between wild-type and knockout mice at both time points.

derived from wild-type or TCCR-deficient mice (data not shown). Second, the effect was reproducible in an APC-free system in which T-cell differentiation was carried out using plate-immobilized anti-CD3/CD28 (Fig. 5b). The reduction in IFN- γ production also correlates with a decrease in the number of cells producing IFN- γ as measured by intracellular fluorescence-activated cell sorting (FACS) staining (data not shown). The observed Th1 deficiency did not appear to be the result of a defect in the IL-12 receptor, as both subunits of the receptor were expressed normally in activated T-cells. As IL-12 could still promote the proliferation of ConA stimulated T cells from wild-type and TCCR-deficient mice, there seems to be no defect in the IL-12 signalling pathway in these mice (Fig. 5c, d).

We have shown the existence of a new type I cytokine receptor, TCCR, involved in the regulation of Th1 immune responses. Previously, IL-12 and IL-18 were identified as potent stimulators of Th1 responses; however, mice deficient in these factors are not completely devoid of Th1 cells, suggesting that additional factors are involved²². Also, it has been suggested that IL-12 signalling alone is not sufficient to drive IFN- γ production in Th2 cells^{23,24}. Despite the similarities in phenotype between our mice and mice deficient in IL-12, IL-12R β 1 or STAT4, as well as the homology between TCCR and the IL-12R β 2 chain, TCCR does not appear to be part of the IL-12 receptor complex. First, a fully functional IL-12 receptor can be reconstituted with the IL-12R β 1 and IL-12R β 2 chains²⁵; and second, TCCR does not bind to IL-12 or specifically associate with the IL-12R chains (data not shown). Furthermore, TCCR-deficient

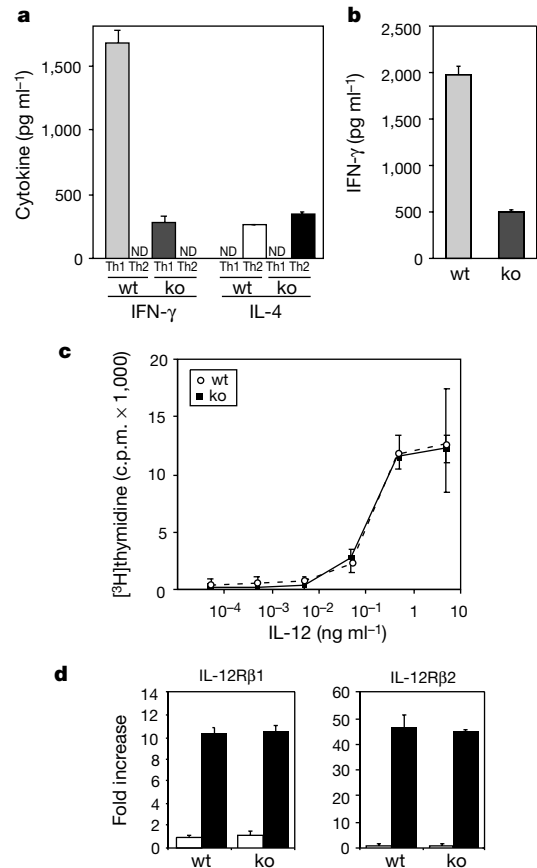


Figure 5 *In vitro* induction of T-helper cell differentiation and proliferation. **a**, **b**, CD4⁺ T cells purified from the spleens of wild-type or TCCR-deficient mice were differentiated into Th1 or Th2 cells (see Methods) in the presence of ConA and irradiated wild-type APC (**a**), or with anti-CD3 and anti-CD28 as stimuli (**b**). Production of IFN- γ and IL-4 was determined by ELISA. Data represent the mean $\pm$ s.d. of pools of five mice per group. ND, not detected. **c**, IL-12 induced proliferation of splenocytes from wild-type and TCCR-deficient mice. ConA-activated splenocytes were incubated for 24 h in the presence of increasing concentrations of IL-12 as indicated. Proliferation of cells was measured by incorporation of [³H]thymidine during the final 6 h (see Methods). **d**, IL-12R mRNA levels in unstimulated (white bars) and ConA-stimulated (black bars) splenocytes. Splenic T cells were stimulated with ConA for 72 h, and mRNA levels for IL-12R β 1 and IL-12R β 2 were determined by real-time quantitative PCR (Taqman; see Methods). Fold increases are relative to the levels of RNA present in wild-type unstimulated cells.

lymphocytes, which express normal levels of both IL-12R chains, proliferate in response to IL-12 stimulation. Thus, it is likely that the TCCR-ligand is different from IL-12. The reduction in IFN- γ production is observed even in the absence of APC, suggesting that activated T cells may be an autocrine source of TCCR ligand, which would then synergize with APC-derived IL-12 for the development of an optimal Th1 response. However, we can not rule out that the TCCR ligand might be produced by another cell type and be present in the serum used in our cultures. In essence, our data provide novel insights into the mechanisms of Th1 differentiation. The apparent specificity of the phenotype described here makes TCCR and its potential ligand candidate targets for therapeutic intervention in Th1-mediated autoimmune disease and allograft rejection. □

Methods

Generation of TCCR knockout mice

A targeting vector was constructed using a 2.4-kb *XhoI*–*HindIII* fragment containing the first 2 exons and a 6.0-kb *EcoRI*–*BamHI* fragment containing exons 9–14. Linearized DNA (20 μ g) was used to electroporate 1×10^7 EGS²⁰ cells. G418/Gancyclovir resistant

clones were screened for homologous recombination by Southern analysis. Genomic DNA was digested with *Bam*HI and hybridized with a 700-base-pair fragment located upstream of the 2.4-kb *Xho*I–*Hind*III arm. Targeted clones were used to generate chimaeric animals by injection into C57BL/6 blastocysts. We bred male chimaeras with C57BL/6 females and screened tail biopsies of agouti-coloured offspring for the TCCR-knockout allele. The TCCR-knockout allele was transferred into inbred C57BL/6 strain by backcrossing heterozygous mice with C57BL/6 mice. The mice were analysed with 58 polymorphic markers dispersed in 25-cM intervals across all autosomes²⁶. After seven generations, there was a greater than 98% contribution of C57BL/6 markers. Only one marker located near the TCCR locus on chromosome 8 was still of the 129 genotype.

T-cell responses

For anti-KLH responses, we immunized mice with 100 µg KLH in saline, in a 1:1 emulsion with complete Freund's adjuvant (CFA), containing 1 mg ml⁻¹ *Mycobacterium tuberculosis* strain H37Ra, (Difco Laboratories, Detroit, MI) in the hind footpads. After 9 d we removed popliteal lymph nodes and prepared cell suspensions. The lymph-node cells were cultured (5 × 10⁵ per well) in various concentrations of KLH in DMEM supplemented with 5% FCS. Proliferation was measured by addition of 1 µCi of [³H]thymidine (ICN, Irvine, CA) for the last 18 h of a 5-d culture, and incorporation of radioactivity was assayed by liquid scintillation counting. We carried out assays for cytokine production by T cells by culturing 5 × 10⁵ draining lymph-node cells either from KLH-primed wild-type or from TCCR-deficient mice in the presence of indicated amounts of the KLH in 96-well plates in a final volume of 200 µl. After 96 h of culture, 150 µl of culture supernatant was removed from each well and cytokine levels were determined by enzyme-linked immunosorbent assay (ELISA) using antibodies from Pharmingen (San Diego, CA), in the recommended conditions.

In vitro induction of T-cell differentiation

We purified CD4⁺ T cells from spleen and lymph nodes of wild-type or TCCR-deficient littermates with anti-CD4 magnetic beads (MACS). Purified T cells (10⁶ cells per ml) were activated in the presence of irradiated (3,000 rad) syngeneic wild-type or knockout APCs (10⁶ per ml) and ConA (2.5 µg ml⁻¹, Boehringer, Mannheim, Germany), or by plating on plates coated with 5 µg ml⁻¹ anti-CD3 and 1 µg ml⁻¹ anti-CD28 antibodies. The culture medium was supplemented with IL-2 (20 U ml⁻¹), IL-12 (3.5 ng ml⁻¹, R&D Systems) and 500 ng ml⁻¹ anti-IL-4 antibody (Pharmingen) for Th1 differentiation, and with IL-2 (20 U ml⁻¹), IL-4 (10³ U ml⁻¹, R&D Systems) and 500 ng ml⁻¹ of anti-IFN-γ antibody (Pharmingen) for Th2 differentiation. After 3 d cells were either lysed for RNA extraction, or extensively washed, counted and restimulated at 10⁶ cells per ml, either in the presence of ConA (2.5 µg ml⁻¹) or on plates coated with 5 µg ml⁻¹ anti-CD3 antibody. After 24 h supernatants were collected and analysed for the presence of cytokines.

Total and OVA-specific immunoglobulin levels

We bled unimmunized mice at 12 weeks of age or older and analysed serum for various isotypes of immunoglobulins by ELISA. For anti-OVA specific antibodies, 6-week-old wild-type or TCCR-deficient mice were immunized with 100 µg of OVA in CFA (intraperitoneally (i.p.)), and challenged 21 d later with 100 µg of OVA in incomplete Freund's adjuvant (i.p.). Seven days after challenge mice were bled and serum was analysed for presence of OVA-specific antibodies.

Real-time PCR analysis

Murine splenocytes were separated into T-helper cells (CD4⁺, F4/80 negative, 97% pure), B cells (CD19⁺, 97% pure), natural killer cells (NK1.1⁺, 99% pure) and macrophages (F4/80 positive, >95% pure) by FACS, and into cytotoxic T cells (CD8⁺, 85% pure) by MACS. Total RNA was extracted with Qiagen RNeasy columns and digested with DNase I to remove contaminating DNA. RNA was probed for TCCR using Taqman¹⁸. All reactions were made in duplicates and normalized to rpl19, a ribosomal housekeeping gene. A 'no-reverse-transcription' control reaction was included and showed that all samples were free of contaminating DNA. The sequence of all primers and probes is available on request.

Received 11 May; accepted 21 August 2000.

- Mosmann, T. R. & Coffman, R. L. Heterogeneity of cytokine secretion patterns and functions of helper T cells. *Adv. Immunol.* **46**, 111–147 (1989).
- Mosmann, T. R. & Sad, S. The expanding universe of T-cell subsets: Th1, Th2 and more. *Immunol. Today* **17**, 138–146 (1996).
- Constant, S. L. & Bottomly, K. Induction of Th1 and Th2 CD4⁺ T cell responses: the alternative approaches. *Annu. Rev. Immunol.* **15**, 297–322 (1997).
- Bazan, J. F. Structural design and molecular evolution of a cytokine receptor superfamily. *Proc. Natl Acad. Sci. USA* **87**, 6934–6938 (1990).
- O'Garra, A. Cytokines induce the development of functionally heterogeneous T-helper cell subsets. *Immunity* **8**, 275–283 (1998).
- Hsieh, C. S. et al. Development of TH1 CD4⁺ T cells through IL-12 produced by Listeria-induced macrophages. *Science* **260**, 547–549 (1993).
- Seder, R. A., Gazzinelli, R., Sher, A. & Paul, W. E. Interleukin 12 acts directly on CD4⁺ T cells to enhance priming for interferon gamma production and diminishes interleukin 4 inhibition of such priming. *Proc. Natl Acad. Sci. USA* **90**, 10188–10192 (1993).
- Le Gros, G. et al. Generation of interleukin 4 (IL-4)-producing cells *in vivo* and *in vitro*: IL-2 and IL-4 are required for *in vitro* generation of IL-4-producing cells. *J. Exp. Med.* **172**, 921–929 (1990).
- Swain, S. L. Helper T-cell subsets: phenotype, function and the role of lymphokines in regulating their development. *Immunol. Rev.* **123**, 115–144 (1991).
- Kuhn, R., Rajewsky, K. & Muller, W. Generation and analysis of interleukin-4 deficient mice. *Science* **254**, 707–710 (1991).

- Noben-Trauth, N. et al. An interleukin 4 (IL-4)-independent pathway for CD4⁺ T cell IL-4 production is revealed in IL-4 receptor-deficient mice. *Proc. Natl Acad. Sci. USA* **94**, 10838–10843 (1997).
- Shimoda, K. et al. Lack of IL-4-induced Th2 response and IgE class switching in mice with disrupted Stat6 gene. *Nature* **380**, 630–633 (1996).
- Takeda, K. et al. Essential role of Stat6 in IL-4 signalling. *Nature* **380**, 627–630 (1996).
- Kaplan, M. H., Schindler, U., Smiley, S. T. & Grusby, M. J. Stat6 is required for mediating responses to IL-4 and for development of Th2 cells. *Immunity* **4**, 313–319 (1996).
- Magrath, J. et al. IL-12-deficient mice are defective in IFN-γ production and type 1 cytokine responses. *Immunity* **4**, 471–481 (1996).
- Wu, C., Ferrante, J., Gately, M. K. & Magrath, J. Characterization of IL-12 receptor β1 chain (IL-12Rβ1)-deficient mice: IL-12Rβ1 is an essential component of the functional mouse IL-12 receptor. *J. Immunol.* **159**, 1658–1665 (1997).
- Kaplan, M. H., Sun, Y. L., Hoey, T. & Grusby, M. J. Impaired IL-12 responses and enhanced development of Th2 cells in Stat4-deficient mice. *Nature* **382**, 174–177 (1996).
- Heid, C. A., Stevens, J., Livak, K. J. & Williams, P. M. Real time quantitative PCR. *Genome Res.* **6**, 986–994 (1996).
- Snapper, C. M. & Paul, W. E. Interferon-γ and B cell stimulatory factor-1 reciprocally regulate Ig isotype production. *Science* **236**, 944–947 (1987).
- Huang, S. et al. Immune response in mice that lack the interferon-γ receptor. *Science* **259**, 1742–1745 (1993).
- Yang, D. D. et al. Differentiation of CD4⁺ T cells to Th1 cells requires MAP kinase JNK2. *Immunity* **9**, 575–585 (1998).
- Takeda, K. et al. Defective NK cell activity and Th1 response in IL-18-deficient mice. *Immunity* **8**, 383–390 (1998).
- Heath, V. L. et al. Cutting edge: ectopic expression of the IL-12 receptor β2 in developing and committed Th2 cells does not affect the production of IL-4 or induce the production of IFN-γ. *J. Immunol.* **164**, 2861–2865 (2000).
- Nishikomori, R., Ehrhardt, R. O. & Strober, W. T-helper type 2 cell differentiation occurs in the presence of interleukin 12 receptor β2 chain expression and signaling. *J. Exp. Med.* **191**, 847–858 (2000).
- Presky, D. H. et al. A functional interleukin 12 receptor complex is composed of two β-type cytokine receptor subunits. *Proc. Natl Acad. Sci. USA* **93**, 14002–14007 (1996).
- Dietrich, W. F. et al. A comprehensive genetic map of the mouse genome. *Nature* **380**, 149–152 (1996).

Supplementary information is available on Nature's World-Wide Web site (<http://www.nature.com>) or as paper copy from the London editorial office of Nature.

Acknowledgements

We thank the Genentech micro-injection lab, DNA sequencing lab and M. Vasser and the DNA synthesis lab for their support; J. Tepper, S. Fong, M. Aguet and R. Schreiber for their advice on this project; and J. Ligos for help with graphics.

Correspondence and requests for materials should be addressed to F.d.S. (e-mail: sauvage@gene.com). The human and murine TCCR sequences have been deposited at GenBank under accession numbers AF265242 and AF265243, respectively.

The nuclear receptor CAR mediates specific xenobiotic induction of drug metabolism

Ping Wei, Jun Zhang, Margarete Egan-Hafley, Shuguang Liang & David D. Moore

Department of Molecular and Cellular Biology, Baylor College of Medicine, 1 Baylor Plaza, Houston, Texas 77030, USA

Organisms encounter a wide range of foreign compounds—or 'xenobiotics'—with potentially harmful consequences. The cytochrome P450 (CYP) enzymes metabolize xenobiotics and thus are a primary defence against these compounds. Increased expression of specific CYP genes in response to particular xenobiotics is a central component of this defence¹, although such induction can also increase production of toxic metabolites. Here we show that the nuclear receptor CAR mediates the response evoked by a class of xenobiotics known as the 'phenobarbital-like inducers'. The strong activation of *Cyp2b10* gene expression by phenobarbital, or by the more potent TCPOBOP, is absent in mice lacking the CAR gene. These animals also show decreased metabolism of the classic CYP substrate zoxazolamine and a complete loss of the

liver hypertrophic and hyperplastic responses to these inducers. Cocaine causes acute hepatotoxicity in wild-type mice previously exposed to phenobarbital-like inducers and this toxicity is also absent in the CAR-deficient animals. Thus, loss of CAR function alters sensitivity to toxins, increasing or decreasing it depending on the compound. Modulation of CAR activity in humans may significantly affect metabolism of drugs and other xenobiotics.

Phenobarbital (PB) exemplifies a chemically diverse set of compounds known as 'PB-like inducers'. Exposure of animals to any of these compounds results in a potent activation of expression of a specific subset of CYP enzymes and other proteins associated with xenobiotic metabolism¹. In the mouse, PB-like inducers increase expression of *Cyp2b10* and several other genes. The pesticide contaminant 1,4-bis[2-(3,5-dichloropyridyloxy)]benzene, referred to as TCPOBOP (ref. 2), is generally considered to be the most potent of this group of inducers. Recently, the nuclear hormone receptor CAR (NR1I3) was implicated in the response to PB-like inducers: CAR activates expression of murine *Cyp2b10* by binding to a phenobarbital response element identified upstream of the gene^{3,4}. This activity can be increased by TCPOBOP, which is an agonist ligand for CAR⁵. However, PB is a better agonist ligand for the related receptor SXR/PXR than for CAR^{6,7}.

To assess the functional role of CAR, we generated two independent mouse lines in which the coding region for β -galactosidase replaces a segment of the CAR gene that is next to the promoter and includes part of the DNA-binding domain (Fig. 1a). As expected, these β -galactosidase knock-in animals are unable to express CAR messenger RNA (Fig. 1c). This loss of CAR expression does not result in any overt phenotype; homozygous CAR^{-/-} animals (Fig. 1b) are born at expected mendelian frequency and both male and female CAR deficient animals are fertile.

CAR has previously been reported to be predominantly expressed in liver^{8,9}. To define the pattern of CAR expression in more detail, β -galactosidase expression was examined in CAR^{+/-} heterozygotes. As expected, the β -galactosidase marker was expressed in liver, and expression was also observed in small intestine. A more detailed examination of the pattern of CAR expression will be described elsewhere.

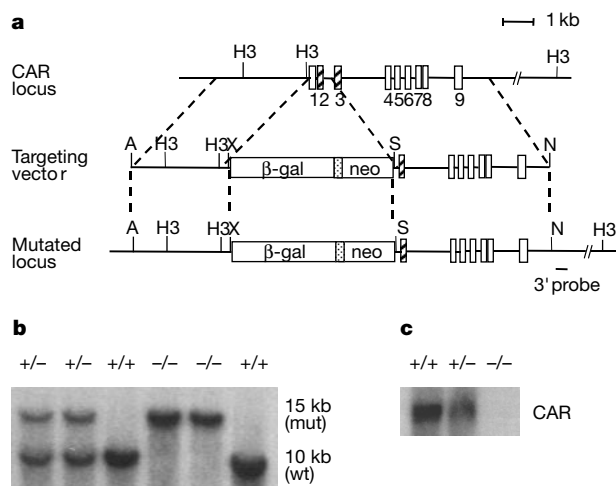


Figure 1 Targeted disruption of the mouse CAR gene. **a**, Diagram of the murine CAR gene. Boxes represent exons. Exons 2 and 3 (hatched boxes) contain the DNA-binding domain. Homologous recombination resulted in replacement of exons 1 and 2 with the β -gal and *neo* resistance genes. Restriction enzyme sites are indicated: A, *Apal*; H3, *HindIII*; N, *NotI*; S, *SalI*; X, *XbaI*. **b**, Genotype analysis by Southern blotting. Genomic DNA from tail samples was digested with *HindIII* and hybridized with the 3' probes. 10-kb or 15-kb bands were generated from wild-type and mutant alleles, respectively. **c**, Northern blot analysis. The murine CAR cDNA was used as the probe to determine the level of CAR mRNA expressed in liver of wild-type, CAR^{+/-} or CAR^{-/-} animals.

Table 1 Zoxazolamine paralysis test

	CAR ^{+/+}	CAR ^{-/-}
Control	1/6 6 h paralysis 4/6 > 12 h paralysis; recovered 1/6 > 12 h paralysis; dead	5/5 > 12 h paralysis; dead
PB	not paralysed	5/6 > 12 h paralysis; dead 1/6 > 12 h paralysis; recovered
TCPOBOP	not paralysed	4/5 > 12 h paralysis; dead 1/5 > 12 h paralysis; recovered

Three days after pretreatment with PB or TCPOBOP, female mice (8–10 weeks old) were given a single intraperitoneal injection of zoxazolamine (300 mg per kg body weight). Paralysis duration was recorded as the time when the mice were able to right themselves repeatedly. Very similar results were obtained with male animals treated with 225 mg per kg body weight (3–5 animals per group), except that the duration of paralysis of the untreated control animals was about 2 h, and none of the knockout animals recovered after more than 12 h of paralysis.

To examine the involvement of CAR in the response to PB-like inducers, we treated wild-type and CAR^{-/-} animals with either PB or TCPOBOP. The robust induction of expression of *Cyp2b10* mRNA in response to either compound in wild-type male or female animals is completely absent in the knockout animals (Fig. 2). Analysis of polyA⁺ mRNA indicates that basal expression of *Cyp2b10*, although very low in the wild-type animals, is also decreased in CAR^{-/-} homozygotes. We observed a less marked induction of CAR in wild-type small intestine but not in CAR^{-/-} small intestine (data not shown).

Acute treatments with PB-like inducers, particularly TCPOBOP, cause an increase in liver mass up to 2-fold relative to total body mass¹⁰. This hepatomegaly is thought to be a reflection of both cellular hypertrophy¹¹ and mitogenesis¹². The CAR^{-/-} animals showed no increase in liver mass after three days of treatment with either PB or TCPOBOP (Fig. 3a). The xenobiotic induction of DNA synthesis revealed by increased incorporation of 5-bromodeoxyuridine (BrdU) observed in wild-type animals is also completely absent in the CAR^{-/-} animals (Fig. 3b).

These results demonstrate that CAR is essential for responses to PB-like inducers. This conclusion was confirmed and extended by examination of the effect of the loss of CAR expression on metabolism of two xenobiotics. The first is the classic substrate zoxazolamine¹³. Increased CYP enzyme activity results in increased metabolic inactivation of this muscle relaxant, as reflected by decreased duration of zoxazolamine-induced paralysis. Most wild-type animals treated with zoxazolamine recovered after more than 12 h of paralysis, while animals pretreated with either PB or TCPOBOP were not paralysed (Table 1). Essentially all of the CAR-deficient animals were also paralysed for more than 12 h, but did not recover, even if pretreated with PB or TCPOBOP. Thus, the knockout animals have dramatically decreased response to either inducer, in agreement with the results above. The mortality of the untreated knockout animals also indicates that loss of CAR function significantly increases zoxazolamine sensitivity.

Treatment with PB-like inducers also sensitizes animals to the hepatotoxic effects of a number of compounds, including cocaine¹⁴.

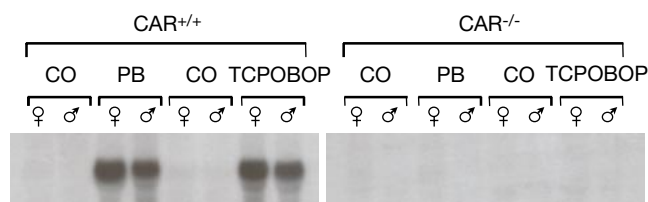


Figure 2 Xenobiotic activation of the *Cyp2b10* gene by CAR. 8-week-old mice were treated with corn oil (CO), PB (24 h) or TCPOBOP (6 h). Total liver RNA was used for northern blot analysis. Three mice were used for each treatment and one representative blot probed with *Cyp2b10* is shown. Similar results were obtained with both independent CAR^{-/-} lines.

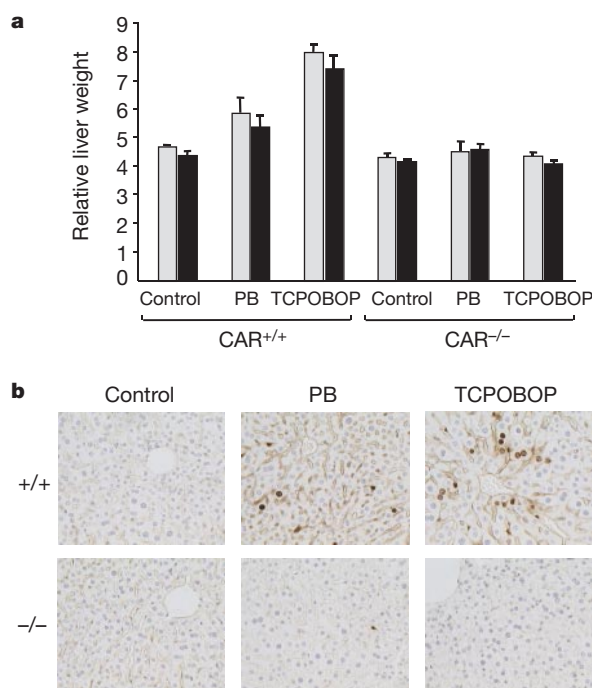


Figure 3 Liver enlargement and hepatocyte proliferation caused by PB or TCPOBOP. **a**, Mice were treated with PB or TCPOBOP for 3 days. Both liver mass and body weight were measured. Percentage of liver mass relative to total body weight is presented as mean $\pm$ s.e.m. determined in at least three individual animals. Grey bars, female; black bars, male. The livers of wild-type PB-treated animals are significantly larger than those of untreated wild-type animals (females, $P = 0.006$; males, $P = 0.014$; Student's t -test) or PB-treated knockout animals (females, $P = 0.0031$; males, $P = 0.0001$). The livers of wild-type TCPOBOP-treated animals are significantly larger than those of untreated wild-type (females, $P = 0.0001$; males, $P < 0.0001$) or TCPOBOP-treated knockout animals (females, $P = 0.0002$; males, $P < 0.0001$). **b**, PB- or TCPOBOP-treated mice were treated with BrdU for 2 h before liver tissues were harvested. Representative microphotography of sections stained with a mouse anti-BrdU monoclonal antibody illustrates the presence of BrdU-positive hepatocytes only in PB- or TCPOBOP-treated wild-type animals.

Treatment with either PB or TCPOBOP results in a very significant increase in serum levels of the liver enzyme alanine aminotransferase (ALT) as an acute response to cocaine administration (Fig. 4). This evidence of liver damage was not observed in CAR^{-/-} animals.

These results clearly demonstrate that CAR is required for the response to PB-like inducers of xenobiotic metabolism. With earlier results demonstrating that CAR can transactivate a phenobarbital response element³⁻⁵, and that the xenobiotic TCPOBOP is a direct CAR agonist⁵, we conclude that CAR functions as a xenobiotic receptor *in vivo* to mediate the response to PB-like inducers. CAR can therefore be added to the peroxisome proliferator activated receptor α and the aryl hydrocarbon receptor as a primary determinant of the response of phase I metabolic enzymes to foreign compounds¹. CAR is joined in this by its closest relative within the receptor superfamily, PXR/SXR, which mediates the response to a distinct group of xenobiotics^{6,15-18}. Although both DNA-binding specificity^{4,5} and xenobiotic responses⁷ of CAR and PXR/SXR overlap to some extent, no evidence for compensatory effects of the latter was observed in the CAR knockout animals. Thus, it is now apparent that specific xenobiotics can induce specific metabolic responses by activating these related but functionally distinct receptors.

Although this can allow increased clearance of potentially deleterious compounds, it can also lead to undesirable drug-drug interactions in which the presence of one drug, such as phenobar-

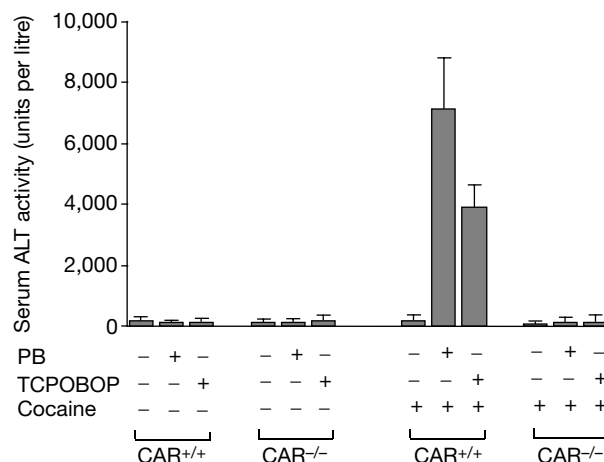


Figure 4 Effect of PB or TCPOBOP on cocaine-mediated hepatotoxicity. Male mice were pretreated with PB or TCPOBOP for 3 days. The animals were injected with cocaine 24 h after the last dose of PB. Blood was drawn 22 h after cocaine treatment for determination of serum ALT activity. Values are means $\pm$ s.e.m. determined in at least three individual animals.

bital, causes increased metabolism of another drug. This can decrease a desirable therapeutic activity or possibly cause acute toxicity through production of harmful metabolites. Differential activation of xenobiotic receptors caused by differences in exposure to specific xenobiotics could also explain the significant inter-individual variability of the levels of particular CYPs. Consistent with the very low basal amounts of mouse Cyp2b10, most humans have low or undetectable amounts of CYP2B6, a human target of CAR activation¹⁹. However, this enzyme is present at up to 100-fold higher levels in a subset of individuals²⁰, potentially as a result of variations in CAR activity. Such variability could be the basis for the relatively rare but clinically significant hepatotoxicity observed in a subset of individuals exposed to high levels of cocaine²¹. More generally, we believe that variations in CAR activity have a significant impact on the metabolism of a wide range of pharmacological agents and other foreign compounds. Defining both the numerous compounds able to activate CAR and the specific downstream target genes that mediate its effects should provide insights into this important xenobiotic response. □

Methods

Targeting vector construction

To construct the targeting vector for the CAR locus, an *Xba*I–*Eag*I fragment containing the nuclear localized β -galactosidase gene from vector pPD 46.21 (obtained from M. J. Tsai) was cloned into the *Xba*I and *Eco*RI sites of the pGKneo plasmid. DNA from AB1 ES cells was used to amplify CAR genomic fragments for both 5' and 3' arms. For the 5' arm, a 3-kilobase (kb) CAR promoter fragment was cloned into the *Ap*aI and *Xba*I sites. For the 3' arm, a 5-kb fragment spanning exons 3 to 9 was cloned into the *S*alI and *N*otI sites. The primers for the 5' arm were 5'-GCGCGCGGGCCCTGCATACATTAA-CACAAACACATACATAT-3' and 5'-GCGCGCTCTAGAAGGACCCAGACTCTGG-ACCCAGGGCAAGA-3'. The primers for the 3' arm were 5'-GCGCGCGTCCGACAGGTGAAGTGCTTCTCCCAACAGAAACAA-3' and 5'-GCGCGCGCGGCCGCTGTCTCTGG GAGCAGCCTCTGCAGCCGCT-3'.

Generation of CAR null mice

AB1 ES cells (10^7) were electroporated with 25 μ g targeting construct in 0.9 ml PBS using a Bio-Rad Gene Pulser (500 μ F; 230 V). The cells were then plated on one or two 10-cm plates containing a monolayer of irradiated STO feeder cells. After 24 h, they were subjected to G418 selection (350 μ g ml⁻¹, Gibco) for 9 days. Resistant clones were analysed by Southern blotting after *H*indIII digestion, using the 3' probe indicated in Fig. 1a. The primers for the 3' probe were 5'-GGACAACCTCAGCCACAGTGATGC-3' and 5'-TCCTTTTGTTACCACTGACTCTGC-3'. Two positive clones were expanded and injected into C57BL/6 blastocysts. Male chimaeras were back-crossed to C57BL/6 females. Heterozygotes were determined by Southern blotting and intercrossed to generate homozygotes.

Animal treatment

At least 3 mice between 8 and 10 weeks old were used for each treatment. Mice were pretreated by intraperitoneal injection with corn oil, PB (100 mg per kg body weight, Sigma) or TCPOBOP (3 mg per kg body weight, a gift from S. Safe) for indicated time. For 3-day PB treatment, mice were injected intraperitoneally three times with PB, one injection per day.

Zoxazolamine paralysis test

Mice pretreated with corn oil, PB or TCPOBOP were given a single intraperitoneal injection of zoxazolamine (300 mg per kg body weight, Sigma), 24 h after the last dose of PB. Mice were placed on their backs and paralysis time was defined as the time required for the animal to regain sufficient consciousness to right itself repeatedly²².

Cocaine treatment and ALT assay

Male mice pretreated with corn oil, PB or TCPOBOP were injected intraperitoneally with cocaine HCl (30 mg per kg body weight), 24 h after the last injection of PB. The mice were anaesthetized 22 h after cocaine treatment. Blood was drawn from the eye for determination of serum alanine aminotransferase (ALT) activity. ALT activity was determined using Vitros ALT slides (Johnson & Johnson Clinical Diagnostics). The procedure was performed at the Methodist Hospital in Houston.

RNA analysis

20 µg of total RNA from individual mouse livers was subjected to northern blot analysis. A mouse CAR complementary DNA probe was used to reveal the absence of CAR transcripts in the CAR null mice. Probes for Cyp2b10 were prepared by polymerase chain reaction after reverse transcription of RNA (RT-PCR) with mouse liver total RNA using Superscript One-step RT-PCR System (Life Technologies). PCR primers were 5'-CCGCCTC TAGAAGTCAACATTGGTTAGAC-3' and 5'-CCGCCGGATCCACACTAAGCCTCAT AAT-3'.

Determination of hepatocyte proliferation following PB or TCPOBOP treatment

Mice pretreated with corn oil, PB or TCPOBOP received a single intraperitoneal dose of BrdU/FdU (2 ml per 100 g body weight, Amersham). Mice were killed 2 h after BrdU administration. BrdU incorporation was determined using a mouse anti-BrdU monoclonal antibody (DAKO) and Vectastain ABC Kit (Vector Laboratories).

Received 5 June; accepted 27 July 2000.

- Waxman, D. J. P450 gene induction by structurally diverse xenochemicals: central role of nuclear receptors CAR, PXR, and PPAR. *Arch. Biochem. Biophys.* **369**, 11–23 (1999).
- Poland, A. *et al.* 1,4-Bis[2-(3,5-dichloropyridyloxy)]benzene, a potent phenobarbital-like inducer of microsomal monooxygenase activity. *Mol. Pharmacol.* **18**, 571–580 (1980).
- Honkakoski, P., Zelko, I., Sueyoshi, T. & Negishi, M. The nuclear orphan receptor CAR-retinoid X receptor heterodimer activates the phenobarbital-responsive enhancer module of the CYP2B gene. *Mol. Cell. Biol.* **18**, 5652–5658 (1998).
- Sueyoshi, T., Kawamoto, T., Zelko, I., Honkakoski, P. & Negishi, M. The repressed nuclear receptor CAR responds to phenobarbital in activating the human CYP2B6 gene. *J. Biol. Chem.* **274**, 6043–6046 (1999).
- Tzamelis, I., Pissios, P., Schuetz, E. G. & D. Moore, D. D. The xenobiotic compound 1,4-bis[2-(3,5-dichloropyridyloxy)]benzene is an agonist ligand for the nuclear receptor CAR. *Mol. Cell. Biol.* **20**, 2951–2958 (2000).
- Jones, S. A. *et al.* The pregnane X receptor: a promiscuous xenobiotic receptor that has diverged during evolution. *Mol. Endocrinol.* **14**, 27–39 (2000).
- Moore, L. B. *et al.* Orphan nuclear receptors, constitutive androstane receptor and pregnane X receptor share xenobiotic and steroid ligands. *J. Biol. Chem.* **275**, 15122–15127 (2000).
- Baes, M. *et al.* A new orphan member of the nuclear receptor superfamily that interacts with a subset of retinoic acid response elements. *Mol. Cell. Biol.* **14**, 1544–1552 (1994).
- Choi, H. S. *et al.* Differential transactivation by two isoforms of the orphan nuclear hormone receptor CAR. *J. Biol. Chem.* **272**, 23565–23571 (1997).
- Heubel, F., Reuter, T. & Gerstner, E. Differences between induction effects of 1,4-bis[2-(3,5-dichloropyridyloxy)]benzene and phenobarbital. *Biochem. Pharmacol.* **38**, 1293–1300 (1989).
- Carthew, P., Edwards, R. E. & Nolan, B. M. The quantitative distinction of hyperplasia from hypertrophy in hepatomegaly induced in the rat liver by phenobarbital. *Toxicol. Sci.* **44**, 46–51 (1998).
- Cunningham, M. L. Role of increased DNA replication in the carcinogenic risk of nonmutagenic chemical carcinogens. *Mutat. Res.* **365**, 59–69 (1996).
- Robinson, J. R. & Nebert, D. W. Genetic expression of aryl hydrocarbon hydroxylase induction. Presence or absence of association with zoxazolamine, diphenylhydantoin, and hexobarbital metabolism. *Mol. Pharmacol.* **10**, 484–493 (1974).
- Bornheim, L. M. Effect of cytochrome P450 inducers on cocaine-mediated hepatotoxicity. *Toxicol. Appl. Pharmacol.* **150**, 158–165 (1998).
- Kliewer, S. A. *et al.* An orphan nuclear receptor activated by pregnanes defines a novel steroid signaling pathway. *Cell* **92**, 73–82 (1998).
- Blumberg, B. *et al.* SXR, a novel steroid and xenobiotic-sensing nuclear receptor. *Genes Dev.* **12**, 3195–3205 (1998).
- Lehmann, J. M. *et al.* The human orphan nuclear receptor PXR is activated by compounds that regulate CYP3A4 gene expression and cause drug interactions. *J. Clin. Invest.* **102**, 1016–1023 (1998).
- Xie, W. *et al.* Humanized xenobiotic response in mice expressing nuclear receptor SXR. *Nature* **406**, 435–439 (2000).
- Shimada, T., Yamazaki, H., Mimura, M., Inui, Y. & Guengerich, F. P. Interindividual variations in human liver cytochrome P-450 enzymes involved in the oxidation of drugs, carcinogens and toxic chemicals: studies with liver microsomes of 30 Japanese and 30 Caucasians. *J. Pharmacol. Exp. Ther.* **270**, 414–423 (1994).

- Code, E. L. *et al.* Human cytochrome P4502B6: interindividual hepatic expression, substrate specificity, and role in procarcinogen activation. *Drug Metab. Dispos.* **25**, 985–993 (1997).
- Selim, K. & Kaplowitz, N. Hepatotoxicity of psychotropic drugs. *Hepatology* **29**, 1347–1351 (1999).
- Liang, H. C. *et al.* Cyp1a2(–/–) null mutant mice develop normally but show deficient drug metabolism. *Proc. Natl Acad. Sci. USA* **93**, 1671–1676 (1996).

Acknowledgements

This work was supported by a grant from NIH to D.D.M. We thank F. DeMayo for help with generating the knockout animals.

Correspondence and requests for materials should be addressed to D.D.M. (e-mail: moore@bcm.tmc.edu).

Structure of a serpin–protease complex shows inhibition by deformation

James A. Huntington, Randy J. Read & Robin W. Carrell

Department of Haematology, University of Cambridge, Wellcome Trust Centre for Molecular Mechanisms in Disease, Cambridge Institute for Medical Research, Hills Road, Cambridge CB2 2XY, UK

The serpins have evolved to be the predominant family of serine-protease inhibitors in man^{1,2}. Their unique mechanism of inhibition involves a profound change in conformation³, although the nature and significance of this change has been controversial. Here we report the crystallographic structure of a typical serpin–protease complex and show the mechanism of inhibition. The conformational change is initiated by reaction of the active serine of the protease with the reactive centre of the serpin. This cleaves the reactive centre, which then moves 71 Å to the opposite pole of the serpin, taking the tethered protease with it. The tight linkage of the two molecules and resulting overlap of their structures does not affect the hyperstable serpin, but causes a surprising 37% loss of structure in the protease. This is induced by the plucking of the serine from its active site, together with breakage of interactions formed during zymogen activation⁴. The disruption of the catalytic site prevents the release of the protease from the complex, and the structural disorder allows its proteolytic destruction^{5,6}. It is this ability of the conformational mechanism to crush as well as inhibit proteases that provides the serpins with their selective advantage.

The irreversibility of inhibition achieved by the serpins has made them the principal inhibitors controlling both intra- and extra-cellular proteolytic pathways. In human plasma, antithrombin controls coagulation, C1-inhibitor controls complement activation, and the inhibitors of plasmin and its activators control fibrinolysis. To determine the structural basis of the serpin mechanism we chose another of the plasma serpins, the archetypal member of the family, α₁-antitrypsin^{7,8}. There have been many unsuccessful attempts over the past 20 years to crystallize the serpin–protease complex. Although the half-life of the complex in isolation is of the order of years, the extreme proteolytic susceptibility of the complex coupled with the high concentrations required for protein crystallization result in a level of heterogeneity incompatible with crystal growth. To overcome these difficulties, we purified the complex from a reaction mixture containing an excess of α₁-antitrypsin over trypsin, and set up crystallization trials at 4 °C. Crystals were obtained within two weeks, and SDS–polyacrylamide gel electrophoresis of an isolated crystal at the time of data collection confirmed it contained only intact complex (see Supplementary Information). The crystal structure at 2.6 Å resolution clearly shows

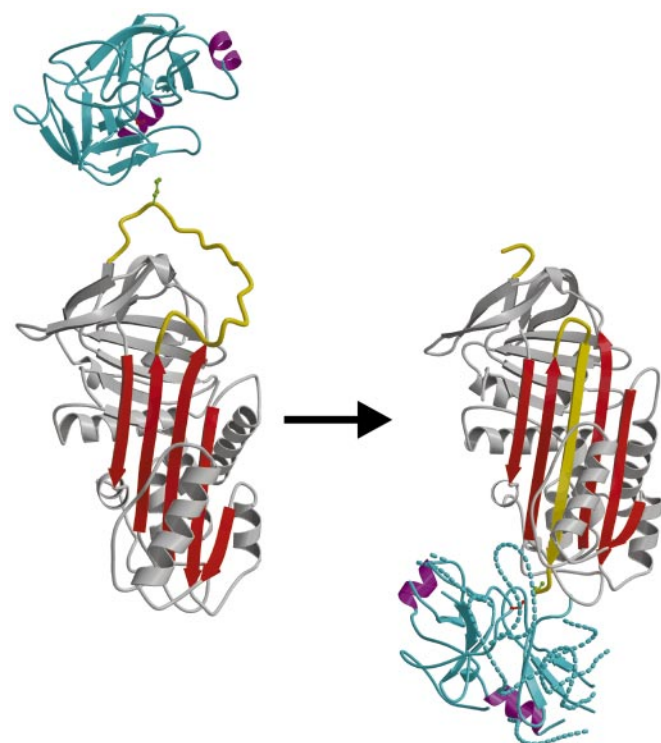


Figure 1 Formation of the complex. Ribbon depictions of native α_1 -antitrypsin⁸ with trypsin aligned above it in the docking orientation (left), and of the complex showing the 71 Å shift of the P1 methionine of α_1 -antitrypsin, with full insertion of the cleaved reactive-centre loop into the A-sheet (right). Regions of disordered structure in the complexed trypsin are shown as interrupted coils projected from the native structure of trypsin. Red, α_1 -antitrypsin β -sheet A; yellow, reactive-centre loop; green ball-and-stick, P1 Met; cyan, trypsin (with helices in magenta for orientation); red ball-and-stick, active serine 195.

the unique mechanism of inhibition used by the serpins. Figure 1 shows the formation of the complex; Fig. 2 shows the distortion of the protease and explains its increased susceptibility to proteolysis; Fig. 3 shows the ester bond between the serpin and the protease, and the disruption of the protease active site. Other figures showing the electron density are available as Supplementary Information.

A glance at the structures in Fig. 1 immediately answers a much-debated question in the field. Wright and Scarsdale⁹ proposed that inhibition involved insertion of the cleaved reactive-centre loop of the serpin into the A β -sheet of the molecule, with a pole-to-pole displacement of the protease. There has however been disagreement as to the extent of loop insertion, with conflicting evidence for both full^{10,11} and partial¹² insertion. Here we see that there is full incorporation of the reactive-centre loop, from its hinge region to the reactive-centre Met 358 (denoted P15–P1). Indeed, the conformation of α_1 -antitrypsin in the complex is precisely superimposable with that of the structure of isolated cleaved α_1 -antitrypsin³ (r.m.s. deviation 0.52 Å for all C α atoms). This structure³ was the starting point for subsequent deductions that the serpins were metastable proteins¹³ with a mobile reactive-centre loop¹⁴ and a spring-like inhibitory mechanism^{9,15}. A more recent serpin structure¹⁶ has shown how initial insertion of the first four residues of the loop (P15–P12) can take place, at which stage the F-helix appears likely to impede further movement of the bulky protease along the sheet. However, Fig. 2 shows how the protease could readily skirt the protuberant F-helix to reach its final position somewhat skew of the central axis of the serpin.

The unexpected finding from the structure of the complex is the degree of conformational disorder induced in the protease. Trypsin is a typical member of the chymotrypsin family and has a well

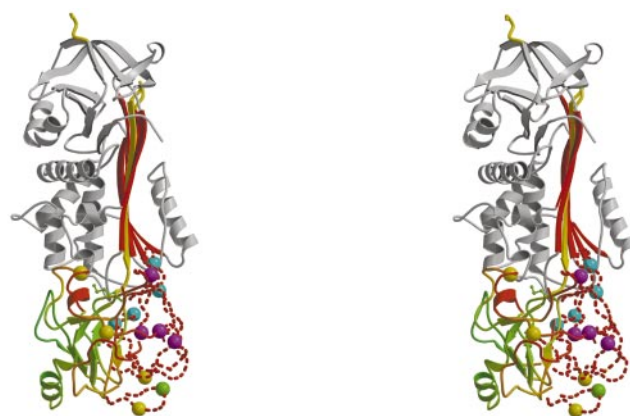


Figure 2 Proteolytic susceptibility of the complexed protease. A stereo side view of the complex coloured according to C α temperature factors for trypsin (α_1 -antitrypsin, coloured as in Fig. 1, retains the low B -factors of its isolated cleaved form). The nine sites of proteolytic cleavage are shown as balls and all occur in regions of crystallographic disorder or high mobility. Cleavage sites: green, of trypsin, by trypsin⁵; yellow, of chymotrypsin, by chymotrypsin⁶; magenta, of chymotrypsin, by neutrophil elastase⁶. Temperature factors from blue to red, going through green at 40 Å², yellow at 60 Å² and red at 90 Å². When the full native trypsin structure is superimposed on the ordered region of trypsin in the complex there are no clashes with symmetry related molecules. The only significant steric overlap is within the asymmetric unit between the serpin and the disordered region of trypsin, as denoted here by cyan balls.

conserved, stable structure that is resistant to proteolysis. That perturbation could occur in the complexed trypsin was predicted from previous studies^{5,6,17} showing a loss of stability and an increase in proteolytic vulnerability of proteases in complexes with serpins. But the surprise is the extent of the disruption observed here, such that although clear and continuous density was seen for more than 60% of the structure of trypsin, some 37% was crystallographically disordered (missing residues 16–41, 62–84, 110–120, 139–156, 186–190, 223–224). The presence within this disordered region of all the previously identified sites of proteolytic cleavage (Fig. 2) is evidence that the disorder represents the changes occurring *in vivo*. The ability to partially denature their cognate proteases is unique to the serpins, and the advantage it gives in making the complex more susceptible to clearance and degradation further explains the biological success of this family. One enzyme that efficiently cleaves complexed proteases⁶ is neutrophil elastase, which is present in high concentrations in inflammatory loci. Such cleavage of the protease in the complex will be of physiological advantage, as it allows localized destruction of the protease before the slower receptor-based uptake of the serpin–protease complex from the circulation.

What is the explanation for the unusual stability of the serpin–protease complex? As compared with the other families of serine protease inhibitors, the complexes of proteases with serpins can persist for months or even years *in vitro*¹⁸. Does the acyl-protease intermediate persist as the result of the exclusion of water, required for hydrolysis, from the active site of trypsin, or is there a distortion of the active site that prevents catalytic deacylation? The structure clearly excludes the first of these possibilities, as the degree of crystallographic disorder, and hence of structural mobility, in the region adjacent to the active site is incompatible with the exclusion of water from the site. Nonetheless, no ordered water molecule is observed in the vicinity of the ester bond or the catalytic His 57. As Fig. 3 shows, there is a gross distortion of the catalytic site of the trypsin with a movement of Ser 195 to a position more than 6 Å away from its catalytic partner His 57. This is well beyond the proximity required for catalytic deacylation, and furthermore the movement of the serine effectively destroys the adjacent oxyanion hole (N–H of Gly 193 and Ser 195) required for the stabilization of

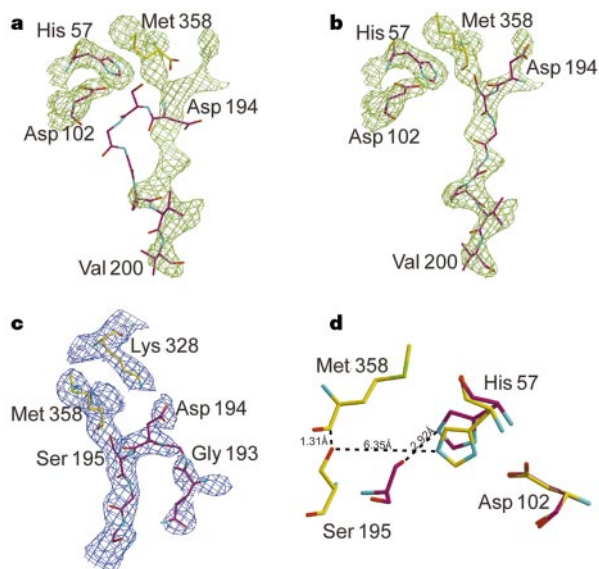


Figure 3 Disruption of active site. The ester bond and distortion of the active site of trypsin is evident from the initial map calculated from the molecular replacement solution of α_1 -antitrypsin (green) shown with the initial (a) and final (b) models of trypsin and P1 Met 358. c, The refined electron density (blue) shows the stretching of the active-site loop resulting in the loss of the oxyanion hole and the replacement of the stabilizing 'activation' salt-bridge between the N-terminal amine and Asp 194 of trypsin with Lys 328 of α_1 -antitrypsin. d, The catalytic triad of native trypsin (magenta) is grossly distorted in the complex (yellow) with a shift of Ser 195 from His 57 well beyond hydrogen-bonding proximity. Superposition is centred on Asp 102.

the tetrahedral transition state. We conclude that this disruption is a direct consequence of the limited length of the serpin reactive-centre loop, which causes the plucking of the ester-linked Ser 195 away from its catalytic partners. The nonspecific nature of this mechanism explains how a single member of the serpin family can inhibit several serine proteases. The serpin–protease interface is limited with only two major contacts expected to contribute to the stability of the complex: Lys 328 of α_1 -antitrypsin forms a salt-bridge with the conserved trypsin Asp 194 (Fig. 3c), and Asn 314 forms three hydrogen bonds with main-chain carbonyl oxygens in trypsin. These interactions are not specific to trypsin and should aid in the stabilization of the complex between α_1 -antitrypsin and any chymotrypsin family member.

The imposed distortion of Ser 195 is directly linked to further distortions and loss of order in the rest of the complexed trypsin molecule. One important contribution is the associated movement of Asp 194 to form a new interaction with Lys 328 of α_1 -antitrypsin (Fig. 3c) while breaking its salt bridge with the free amino group of Ile 16. In trypsinogen, where the amino group of Ile 16 is masked by the propeptide, a region known as the activation domain is disordered but becomes ordered upon activation to trypsin⁴. The activation domain, which comprises residues from the amino terminus to 19, 142–152, 184–193 and 216–223, accounts for about a third of the portion of trypsin that becomes disordered or distorted in the complex with α_1 -antitrypsin. By breaking the interaction involving Asp 194, the serpin reverses the zymogen activation mechanism. Further distortions may be transmitted from Ser 195 through the series of disulphide bridges in trypsin. The disordered region also includes the calcium-binding site. Although each of these perturbations will contribute to the overall disruption of trypsin, the most significant factor is likely to be the steric clash due to the forced overlap with the serpin (Figs 1, 2 and Supplementary Information). The protease is, in effect, crushed against the body of the serpin. This explains why the serpins have evolved so that the energy released upon loop insertion results in a

Table 1 Data processing, refinement and models

Crystals	
Space group	C222 ₁
Cell dimensions (Å)	a = 63.2; b = 171.3; c = 145.8
Solvent content (%)	49.2
Data processing statistics	
Wavelength (Å)	0.87 (Daresbury SRS, station 9.6)
Resolution (Å)	42.3/2.6
Total reflections	190,117
Unique reflections	24,548
$\langle I/\sigma(I) \rangle$	4.3
$\langle \langle I/\sigma(I) \rangle \rangle$	12.0
Completeness (%)	99.2
Multiplicity	7.7
R_{merge}	0.158
Model	
Number of protein/water residues	567/56
Residues modelled	α_1 -antitrypsin trypsin
Average B-factor (Å ²)	43.8
Refinement statistics	
Reflections in working/free set	23,506/1,042
R-factor/ R_{free}	20.5/23.9
r.m.s. deviation of bonds (Å)/angles (°) from ideality	0.0064/1.37
Ramachandran plot; residues in	
most favoured region (%)	85.7
additionally allowed region (%)	13.6
generously allowed region (%)	0.7
disallowed region (%)	0

molecule that is hyperstable, whereas the protease, as with most other proteins, has evolved to be just stable enough to prevent unfolding at ambient temperatures. Thus, the challenge of the steric overlap of the two structures leaves the α_1 -antitrypsin unaffected, but results in a substantial collapse of the ordered structure of trypsin.

The relevance of the mechanism of inhibition presented here to the serpins in general is supported by studies on a variety of serpin–protease complexes using fluorescence, proteolysis and NMR^{5,6,10,11,17–19}. Their collective results outline the changes observed here, but the field has remained unconvinced^{20,21}. With the determination of the crystal structure of the complex these controversies can now be put to rest, but we are also aware that the structure opens further questions and new opportunities for research. One small, but obvious, question has already been answered by an experiment of nature. Is the limited length of the reactive-centre loop of the serpin a critical factor in the distortion of the catalytic triad and hence the stability of the complex? An incidental test of this was reported in a study of a family with a bleeding disorder²². The disorder was caused by the insertion of an extra residue into the reactive-centre loop of the fibrinolysis inhibitor α_2 -antiplasmin. The tightness of the tethering to the serpin reactive-centre loop is therefore a critical factor in causing the distortion of the protease that leads to inhibition. Thus, serpins inhibit serine proteases by a novel mechanism, a fifth example to add to the four recently listed by Bode and Huber²³—inhibition by deformation. □

Methods

The complex of N-terminally histidine-tagged human α_1 -antitrypsin and bovine pancreatic trypsin (Sigma) was formed at pH 7 by incubation for 3 h at room temperature using a 2.9-fold molar excess of α_1 -antitrypsin over trypsin. The reaction was stopped by the addition of 4-(2-aminoethyl)benzenesulfonyl fluoride (Sigma) to 1 mM, and cooling on ice. The complex was purified using a pH gradient from 5.7 to 8.0 on a Poros S column (PerSeptive Biosystems, Framingham, Massachusetts), and concentrated to 10 mg ml^{−1} in 20 mM NaAcetate, pH 5.7. Crystals were obtained from hanging drops containing 0.2 M

tri-NaCitrate and 20% PEG 3350, at a final pH of 7.4 (PEG/Ion Screen, Hampton Research, San Diego, California) within two weeks at 4 °C. Intact complex was verified by SDS–polyacrylamide gel electrophoresis of washed crystals (see Supplementary Information). Data were collected from a single frozen crystal, cryoprotected in 28.5% PEG 4000 and 10% PEG 400, at beamline 9.6 at the SRS Daresbury, UK.

The data were processed using MOSFLM²⁴ and merged using SCALA²⁵ from the CCP4 package²⁶ (Table 1). The molecular replacement solution for α_1 -antitrypsin in the complex was obtained using AMORE²⁷ and the structure of cleaved α_1 -antitrypsin²⁸ as the search model. Conventional molecular replacement searches failed to place a model of intact trypsin²⁹ in the complex, although maps calculated with phases from α_1 -antitrypsin alone showed clear density for the ordered portion of trypsin (Fig. 3 and Supplementary Information). It was immediately apparent that density was only present for about half of the volume expected to be occupied by intact trypsin. A search model comprising trypsin residues 27–124 and 230–245 was orientated using AMORE to compute a domain rotation function³⁰ against structure factors corresponding to a sphere of the ordered density, which were calculated using the program GHKL (L. Tong, unpublished data). The position of the oriented model relative to α_1 -antitrypsin was determined with AMORE using the original diffraction data. The entire model of trypsin was superimposed on the fragment and then truncated to the limits of the electron density to provide an initial model of the complex. The truncated model, to our surprise, was nearly complete in accounting for the ordered structure contributing to the diffraction data, despite including only about 50% of the trypsin residues. In fact, the amount of ordered density changed little throughout the course of refinement. Completeness of this model was estimated at 99% by a σ_A -plot computed in the program SIGMA³¹. The model comprising α_1 -antitrypsin alone was estimated to be 83% complete, whereas α_1 -antitrypsin comprises only 62% of the mass of the complex (σ_A -plots are included as Supplementary Information). The final molecular model was achieved through an iterative procedure of rebuilding using XtalView and refinement in CNS³² using a maximum likelihood target³³. Statistics for data processing, refinement and for the final model are given in Table 1. Figures were made using the programs Molscript³⁴, Bobscript³⁵ and Raster3D³⁶.

Received 10 May; accepted 2 August 2000.

- Potempa, J., Korzus, E. & Travis, J. The serpin superfamily of proteinase inhibitors: structure, function, and regulation. *J. Biol. Chem.* **269**, 15957–15960 (1994).
- Laskowski, M. & Qasim, M. A. What can the structures of enzyme-inhibitor complexes tell us about the structures of enzyme substrate complexes? *Biochim. Biophys. Acta* **1477**, 324–337 (2000).
- Loebermann, H., Tokuku, R., Deisenhofer, J. & Huber, R. Human α_1 -proteinase inhibitor. Crystal structure analysis of two crystal modifications, molecular model and preliminary analysis of the implications for function. *J. Mol. Biol.* **177**, 531–557 (1984).
- Huber, R. & Bode, W. Structural basis of the activation and action of trypsin. *Acc. Chem. Res.* **11**, 114–122 (1978).
- Kaslik, G., Patthy, A., Balint, M. & Graf, L. Trypsin complexed with α_1 -proteinase inhibitor has an increased structural flexibility. *FEBS Lett.* **370**, 179–183 (1995).
- Stavridi, E. S. *et al.* Structural change in α_1 -chymotrypsin induced by complexation with α_1 -antichymotrypsin as seen by enhanced sensitivity to proteolysis. *Biochemistry* **35**, 10608–10615 (1996).
- Huber, R. & Carrell, R. W. Implications of the three-dimensional structure of α_1 -antitrypsin for structure and function of serpins. *Biochemistry* **28**, 8951–8966 (1989).
- Elliott, P. R., Abrahams, J. P. & Lomas, D. A. Wild-type α_1 -antitrypsin is in the canonical inhibitory conformation. *J. Mol. Biol.* **275**, 419–425 (1998).
- Wright, H. T. & Scarsdale, J. N. Structural basis for serpin inhibitor activity. *Proteins* **22**, 210–225 (1995).
- Stratikos, E. & Gettins, P. G. Formation of the covalent serpin–proteinase complex involves translocation of the proteinase by more than 70 Å and full insertion of the reactive center loop into beta-sheet A. *Proc. Natl Acad. Sci. USA* **96**, 4808–4813 (1999).
- Fa, M. *et al.* The structure of a serpin–proteinase complex revealed by intramolecular distance measurements using donor–donor energy migration and mapping of interaction sites. *Structure* **8**, 397–405 (2000).
- Picard, V., Marque, P. E., Paolucci, F., Aiach, M. & Le Bonniec, B. F. Topology of the stable serpin–protease complexes revealed by an autoantibody that fails to react with the monomeric conformers of antithrombin. *J. Biol. Chem.* **274**, 4586–4593 (1999).
- Carrell, R. W. & Owen, M. C. Plasminogen, α_1 -antitrypsin, antithrombin and the mechanism of inflammatory thrombosis. *Nature* **317**, 730–732 (1985).
- Stein, P. E. *et al.* Crystal structure of ovalbumin as a model for the reactive centre of serpins. *Nature* **347**, 99–102 (1990).
- Engh, R. A., Huber, R., Bode, W. & Schulze, A. J. Divining the serpin inhibition mechanism: a suicide substrate ‘spring’? *Trends Biotechnol.* **13**, 503–510 (1995).
- Gooptu, B. *et al.* Inactive conformation of the serpin $\alpha_1(1)$ -antichymotrypsin indicates two-stage insertion of the reactive loop: implications for inhibitory function and conformational disease. *Proc. Natl Acad. Sci. USA* **97**, 67–72 (2000).
- Herve, M. & Ghelis, C. Conformational stability of the covalent complex between elastase and α_1 -proteinase inhibitor. *Arch. Biochem. Biophys.* **285**, 142–146 (1991).
- Olson, S. T. *et al.* Role of the catalytic serine in the interactions of serine proteinases with protein inhibitors of the serpin family. Contribution of a covalent interaction to the binding energy of serpin–proteinase complexes. *J. Biol. Chem.* **270**, 30007–30017 (1995).
- Plotnick, M. I., Mayne, L., Schechter, N. M. & Rubin, H. Distortion of the active site of chymotrypsin complexed with a serpin. *Biochemistry* **35**, 7586–7590 (1996).
- Gils, A. & Declercq, P. J. Structure–function relationships in serpins: current concepts and controversies. *Thromb. Haemost.* **80**, 531–541 (1998).
- Harrop, S. J. *et al.* The crystal structure of plasminogen activator inhibitor 2 at 2.0 Å resolution: implications for serpin function. *Struct. Fold Des.* **7**, 43–54 (1999).
- Holmes, W. E. *et al.* α_2 -antiplasmin Enschede: alanine insertion and abolition of plasmin inhibitory activity. *Science* **238**, 209–211 (1987).
- Bode, W. & Huber, R. Structural basis of the endoproteinase–protein inhibitor interaction. *Biochim. Biophys. Acta* **1477**, 241–252 (2000).
- Leslie A. W. G. in *Joint CCP4 and ESF-EACMB Newsletter on Protein Crystallography* vol. 26

(Daresbury Laboratory, Warrington, 1992).

- Evans P. R. in *Proceedings of the CCP4 Study Weekend. Data Collection and Processing* (eds Sawyer, L., Isaacs, N. & Bailey, S.) 114–122 (Daresbury Laboratory, 1993).
- Collaborative Computational Project Number 4. The CCP4 suite: programs for protein crystallography. *Acta Crystallogr.* **50**, 760–763 (1994).
- Navaza, J. AMORE – An automated package for molecular replacement. *Acta Cryst. A* **50**, 157–163 (1994).
- Engh, R. *et al.* The S variant of human α_1 -antitrypsin, structure and implications for function and metabolism. *Protein Eng.* **2**, 407–415 (1989).
- Lee, S. L. New inhibitors of thrombin and other trypsin-like proteases: hydrogen bonding of an aromatic cyano group with a backbone amide of the P1 binding site replaces binding of a basic side chain. *Biochemistry* **36**, 13180–13186 (1997).
- Colman, P. M., Fehllhammer, H. & Bartles, K. in *Crystallographic Computing Techniques* (eds Ahmed, F. R., Huml, K. & Sedlacek, B.) 248–258 (Munksgaard, Copenhagen, 1976).
- Read, R. J. Improved Fourier coefficients for maps using phases from partial structures with errors. *Acta Cryst. A* **42**, 140–149 (1986).
- Brunger, A. T. *et al.* Crystallography & NMR system: A new software suite for macromolecular structure determination. *Acta Cryst. D* **54**, 905–921 (1998).
- Pannu, N. S. & Read, R. J. Improved structure refinement through maximum likelihood. *Acta Cryst. A* **52**, 659–668 (1996).
- Kraulis, P. J. MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures. *J. Appl. Crystallogr.* **24**, 946–950 (1991).
- Esnouf, R. Further additions to MolScript version 1.4, including reading and contouring of electron-density maps. *Acta Crystallogr. D* **55**, 938–940 (1999).
- Merritt, E. A. & Bacon, D. J. Raster3D: photorealistic molecular graphics. *Methods Enzymol.* **277**, 505–524 (1997).

Supplementary information is available on Nature's World-Wide Web site (<http://www.nature.com>) or as paper copy from the London editorial office of Nature.

Acknowledgements

We thank our colleagues, N. Pannu for advice throughout; D. Lomas for reading the paper; A. Lesk and P. Stein for discussions; and K. Belzar for support. This work was supported by grants from the Wellcome Trust, the European Community and the National Institutes of Health (J.A.H.).

Correspondence and requests for materials should be addressed to J.A.H. (e-mail: jah52@cam.ac.uk) or R.W.C. (e-mail: rwc1000@cam.ac.uk). Atomic coordinates have been deposited in the Protein Data Bank under accession code 1EZX.

errata

Intraprotein radical transfer during photoactivation of DNA photolyase

Corrine Aubert, Marten H. Vos, Paul Mathis, André P. M. Eker & Klaus Brettel

Nature **405**, 586–590 (2000).

Figure 5 of this paper contained an error. The lower right-hand box in the reaction scheme, which read ‘FADH[•] TrpH TrpH TrpH’, should have read ‘FADH[•] TrpH TrpH Trp’.

Neural synchrony correlates with surface segregation rules

Miguel Castelo-Branco, Rainer Goebel, Sergio Neuenschwander & Wolf Singer

Nature **405**, 685–689 (2000).

In Fig. 1b of this Letter, the scale bar for the repetitive fields should be half as large as it was printed. In Fig. 1c, the label that reads PMLS should read A18.

new on the market

programs for LS 6000 scintillation counters. The programs deliver enhanced capabilities for archiving data on a PC or network drive, along with features to aid regulatory compliance. The Data Capture/Network Software module collects scintillation data in the background, and automatically converts Beckman Standard Format files to other formats. The BSF File Utility provides viewing and printing capabilities of GLP/GMP-compliant BSF files, manual file conversion and graphical data representation. The Instrument Performance Tracking module analyses data from LS standards, collected over time, and provides regulatory reports.

Reader Service No. 107

CrystalDiffract 3

CrystalMaker www.crystallmaker.co.uk

Making diffraction easier

CrystalDiffract Version 3.0 is a rewrite of a program developed to handle X-ray or neutron diffraction from crystals. The makers say that this Mac-only program makes powder diffraction intuitive, interactive and even fun. CrystalDiffract simulates X-ray or neutron powder diffraction patterns for any crystal, with controls for manipulating and measuring both calculated and observed diffraction data. A new Parameter List gives users real-time control over the diffraction experiment. The full, professional edition, CrystalDiffract Pro, allows comparison of observed experimental data with simulated data in the same window.

Reader Service No. 108

Mathematica

Wolfram www.wolfram.com

Mathematica gains a new package

Digital Image Processing is a new Mathematica application package with a powerful collection of fundamental and advanced image-processing tools. The package contains a large assortment of image transformation and manipulation tools as well as an interactive manual and tutorial to get new users started quickly.

Reader Service No. 109

Equation Illustrator V

MGCSoft www.mgcsoft.com

Combining graphics and mathematics

MGCSoft has released Equation Illustrator V for Windows 95, 98 and NT4 eases the production of paper documents and web graphics that include math/chemical equations vector graphics and pictures. Equation Illustrator V has been designed to ease the task of combining graphics and complicated formatted text such as mathematical equations in electronic and printed documents. A WYSIWYG interface keeps automation to a minimum. Formatted equation macros can be stored and retrieved with a cut-and-paste interface. Selections can be sent to the Windows clipboard as Enhanced Metafiles and bitmaps or saved to file for use in

www and other documents. Vector-drawing functions are included, from simple rectangles to regular polygons and Bézier curves. Some technology-specific functions such as waves and pulses are also included.

Reader Service No. 110

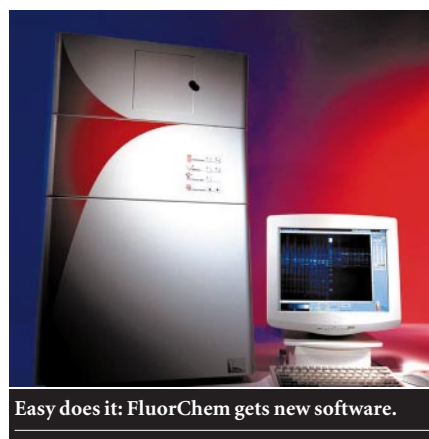
AlphaEase FC 2.0

Alpha Innotech www.alphainnotech.com

Image-analysis software for the FluorChem low-light imaging system

Version 2.0 of the AlphaEase FC image acquisition, analysis and enhancement software, features a new interface designed for multi-user facilities. The software allows for 1 and 2D densitometry, molecular-weight calculations, colony counting, image filtering, image annotation and labelling. New functions include the magic wand facility, which automatically thresholds bands and spots and provides corresponding densitometry information. The 1D multi-lane analysis function has numerous background-subtraction options to include rubber band, minimum profile, rolling disk, and valley to valley. The AutoLane function automatically detects lanes and bands in a gel image with a single button click. The AutoSpot function allows for automatic 2D detection and quantification of spots on an image.

Reader Service No. 111



Easy does it: FluorChem gets new software.

ModelMaker 4.0

ModelKinetix www.cherwell.com

Simulation modelling software

ModelKinetix, part of the Cherwell Scientific group, is aiming ModelMaker 4.0 at the chemical, pharmacokinetics and environmental sciences markets. New features include Monte Carlo analysis, new graphing features and optimization methods, stiff equation solving, the ability to archive model/results, and an improved user-interface for modelling. ModelKinetix say that this program is particularly suited to scientists who are not mathematicians, programmers or modellers.

Reader Service No. 112

These notes are compiled in the Nature office from information provided by the manufacturers.

ADVERTISEMENT

NEW! Reference Manager 9 **FREE TRIAL!**
Search PubMed and 300+ Internet Libraries!

3 RESEARCH TOOLS IN 1
Internet Reference Searcher • Database Manager • Bibliography Builder

"Reference Manager 9 is absolutely essential for my work. Cite While You Write is a must-have feature no researcher should do without."
— Dr. Kuni Suzuki
Director, Neuroscience Center
University of North Carolina

Used by:
Researchers, Librarians, Writers and Students who need to find, import, manage and perfectly format reference data for:
Professional Journals • Curricula Vitae
Regulatory Filings • Research Papers
Dissertations/Theses • Grants

For a FREE trial version of Reference Manager, visit our Web site at:
www.refman.com

Research Information Systems • TOLL FREE 800-722-1227 • phone (703) 436-5035 • fax (703) 438-1073 • E-mail: sales@ris.com
RIS Europe • phone: 44-1835-370616 • fax: 44-1835-259659 • E-mail: uksales@ris.com

READER ENQUIRY NO. 459